Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations | 1. Thesis and Dissertation Collection, all items

1987

An APL workspace for conducting nonparametric statistical inference.

Vagts, Wayne Franz.

http://hdl.handle.net/10945/22195

Downloaded from NPS Archive: Calhoun

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93943-5002

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

AN APL WORKSPACE FOR CONDUCTING
NONPARAMETRIC STATISTICAL INFERENCE

by

Wayne Franz Vagts

June 1987

Thesis Advisor: T. Jayachandran

Approved for public release; distribution is unlimited

T233699

REPORT DOCUMENTATION PAGE

a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
a SECURITY CLASSIFICATION AUTHORITY b DECLASSIFICATION / DOWNGRADING SCHEDULE	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b OFFICE SYMBOL (If applicable) 53	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School
c ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000	7b ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
a NAME OF FUNDING / SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO · PROJECT NO · TASK NO · WORK UNIT ACCESSION NO

TITLE (Include Security Classification)
AN APL WORKSPACE FOR CONDUCTING NONPARAMETRIC STATISTICAL INFERENCE

2 PERSONAL AUTHOR(S) VAGTS, Wayne Franz

3a TYPE OF REPORT Master's Thesis	13b TIME COVERED FROM ______ TO ______	14 DATE OF REPORT (Year, Month, Day) 1987 June	15. PAGE COUNT 116

5 SUPPLEMENTARY NOTATION

COSATI CODES		
FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Nonparametric, APL, Sign Test, Wilcoxon Text, Mann-Whitney Test, Kendall Test, Spearman Test, Nonparametric simple regression, (continued)

9 ABSTRACT (Continue on reverse if necessary and identify by block number)

This thesis contains programs written in APL and documentation for performing certain nonparametric tests and computing nonparametric confidence intervals. These methods of inference are particularly useful in dealing with Department of Defense related problems as illustrated in the several military examples worked in Appendix C. The following nonparametric tests are considered: Sign Test, Wilcoxon Signed-rank Test, Mann-Whitney Test, Kruskal-Wallis Test, Kendall's B, Spearman's R, and Nonparametric Linear Regression. The tests are based on the exact distributions of the respective test statistics unless a large sample approximation is determined to provide at least a three decimal place accuracy. The software consists of two APL workspaces; one, which is designed for microcomputers (IBM PC's or compatibles) and is menu driven,

0 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
a NAME OF RESPONSIBLE INDIVIDUAL T. Jayachandran	22b TELEPHONE (Include Area Code) 408-646-2600 22c OFFICE SYMBOL 53Jy

D FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

Block 19. Abstract (Continued)

and the other, without menus, is designed for the mainframe computer (IBM 3033) at the Naval Postgraduate School.

Block 18. Subject Terms (continued)

Exact Distribution, Asymptotic Approximations.

Approved for public release; distribution is unlimited.

An APL Workspace for Conducting Nonparametric Statistical Inference

by

Wayne Franz Vagts
Lieutenant Commander, United States Navy
B.S., University of Notre Dame, 1975

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 1987

ABSTRACT

This thesis contains programs written in APL and documentation for performing certain nonparametric tests and computing nonparametric confidence intervals. These methods of inference are particularly useful in dealing with Department of Defense related problems as illustrated in the several military examples worked in Appendix C. The following nonparametric tests are considered: Sign Test, Wilcoxon Signed-rank Test, Mann-Whitney Test, Kruskal-Wallis Test, Kendall's B, Spearman's R, and Nonparametric Linear Regression. The tests are based on the exact distributions of the respective test statistics unless a large sample approximation is determined to provide at least a three decimal place accuracy. The software consists of two APL workspaces; one, which is designed for microcomputers (IBM PC's or compatibles) and is menu-driven, and the other, without menus, is designed for the mainframe computer (IBM 3033) at the Naval Postgraduate School.

TABLE OF CONTENTS

LIST OF TABLES....................7

ACKNOWLEDGEMENTS....................8

I. INTRODUCTION....................9

II. WORKSPACE DESIGN ISSUES....................11

III. GENERAL SAMPLE SIZE CONSIDERATIONS AND ASYMPTOTIC APPROXIMATIONS....................13

IV. TESTS FOR LOCATION BASED ON SINGLE AND PAIRED-SAMPLE DATA....................15

- A. ORDINARY SIGN TEST....................15
- B. WILCOXON SIGNED-RANK TEST....................18

V. TESTS BASED ON TWO OR MORE SAMPLES....................24

- A. MANN-WHITNEY TEST....................24
- B. KRUSKAL-WALLIS TEST....................30

VI. TESTS FOR ASSOCIATION IN PAIRED-SAMPLES....................35

- A. KENDALL'S B....................35
- B. SPEARMAN'S R....................38

VII. NONPARAMETRIC SIMPLE LINEAR REGRESSION....................43

- A. COMPUTATION OF THE ESTIMATED REGRESSION EQUATION....................43
- B. HYPOTHESIS TESTING....................44
- C. CONFIDENCE INTERVAL ESTIMATION....................44

VIII. AREAS FOR FURTHER WORK....................46

LIST OF REFERENCES....................47

APPENDIX A: DOCUMENTATION FOR THE MICROCOMPUTER WORKSPACE....................49

APPENDIX B: DOCUMENTATION FOR THE MAINFRAME COMPUTER WORKSPACE............52

APPENDIX C: WORKSPACE FAMILIARIZATION THROUGH PRACTICAL EXAMPLES..............54

APPENDIX D: MAIN PROGRAM LISTINGS FOR MICROCOMPUTER WORKSPACE............78

APPENDIX E: MAIN PROGRAM LISTINGS FOR MAINFRAME COMPUTER WORKSPACE........88

APPENDIX F: LISTINGS OF SUBPROGRAMS BASIC TO BOTH WORKSPACES...............98

APPENDIX G: LISTINGS OF PROGRAMS USED TO GENERATE C.D.F. COMPARISON TABLES......107

INITIAL DISTRIBUTION LIST...........................115

LIST OF TABLES

1. C.D.F. COMPARISONS FOR THE SIGN TEST............16
2. C.D.F. COMPARISONS FOR THE WILCOXON SIGNED-RANK TEST...................20
3. C.D.F. COMPARISONS FOR THE MANN-WHITNEY TEST.........................26
4. C.D.F. COMPARISONS FOR THE KRUSKAL-WALLIS TEST...........................32
5. C.D.F. COMPARISONS FOR THE KRUSKAL-WALLIS TEST USING COMPUTER SIMULATION...........33
6. C.D.F. COMPARISONS FOR KENDALL'S B..............37
7. C.D.F. COMPARISONS FOR SPEARMAN'S R.............41

ACKNOWLEDGMENTS

Many thanks to Professor Larson for his APL workspace STATDIST from which several of the normal theory based asymptotic approximations are computed. APL*PLUS/PC System software and APL*PLUS/PC TOOLS are used in the construction of the microcomputer workspace.[1] IBM's VSAPL is the APL version used for the mainframe workspace.

---

[1] APL*PLUS is a copyrighted software from STSC, Inc., a CONTEL Company, 2115 East Jefferson Street, Rockville, Maryland 20852.

I. INTRODUCTION

Although nonparametric procedures are powerful tools to the analyst, they are currently underused and often avoided by potential users. Perhaps one reason for this is the difficulty in generating the exact distributions of the test statistics, even for moderate sample sizes. Consequently, tables of these distributions are only available for very small sample sizes and normal theory based approximations must then be used.

The purpose of this thesis is to make a variety of nonparametric procedures quick, easy and accurate to apply using menu driven computer programs in APL.[1] These programs use enumeration, recursion, or combinatorial formulas to generate the exact null distribution of the various nonparametric test statistics. This allows hypothesis testing and confidence interval estimation to be based on exact distributions without the use of tables. For larger sample sizes, the normal, F, and T distributions are used to approximate the distributions of the test statistics with three decimal place accuracy.

---

[1] APL was chosen because it is an interactive language that is especially powerful at performing calculations dealing with rank order statistics and vector arithmetic. Menus are not included in the workspace designed for the mainframe.

Section II addresses workspace design issues, to include, workspace requirements and assumptions regarding its use. Section III discusses the methods used to assess the accuracy of different asymptotic approximations, and the sample sizes required for an approximation to yield three decimal place accuracy. Section IV gives background information and discusses programming methodology for nonparametric tests based on single and paired sample data. In Section V, nonparametric tests for two or more independent samples are considered. Section VI discusses nonparametric tests for association; and, Section VII deals with nonparametric simple linear regression. Section VIII recommends other nonparametric tests that may be added to the workspace and areas for further work.

To show application of nonparametric statistical methods to Department of Defense problems, several military examples are worked in Appendix C.

II. WORKSPACE DESIGN ISSUES

This section presents a brief overview of the design considerations used in developing the APL workspace for both the mainframe and microcomputer.

A. EQUIPMENT AND SOFTWARE REQUIREMENTS

The microcomputer must be an IBM PC or AT compatible, equipped with 512 kilobytes of RAM and the APL*PLUS/PC system software, release 3.0 or later, and IBM's DOS, version 2.00 or later.[1] The 8087 math coprocessor chip is not required to run this software, but will increase the computational speed.

B. KNOWLEDGE LEVEL OF THE USER

The user is expected to have had some exposure to APL and a working knowledge of nonparametric statistics. Familiarity with microcomputers or the Naval Postgraduate School mainframe computer is assumed.

---

[1] The APL system software requires 144 kilobytes of RAM while the NONPAR workspace requires an additional 190 kilobytes.

C. SELECTION OF TESTS

The nonparametric tests chosen for this workspace are some of the more widely known, and are considered basic material for any nonparametric statistics course. More information about the tests can be found in any of the textbooks that are referred to in this document.

D. MENU DISPLAYS

The microcomputer's workspace is designed around the use of menus. This was accomplished using the software package PC TOOLS from STSC. These menus are designed to guide the user through the selection of the tests without an excessive amount of prompting. The main menu displays the choices available in the workspace, while the test menus give the background information and options available for each test. Help menus to provide additional information about the tests are also available.

E. ORGANIZATION OF WORKSPACE DOCUMENTATION

Separate documentation is included for the microcomputer's and mainframe computer's workspaces (see Appendices A and B, respectively). These appendices explain the organization and operation of the workspaces. Appendix C, which provides example problems for each nonparametric test, is applicable to both workspaces.

III. GENERAL SAMPLE SIZE CONSIDERATIONS AND ASYMPTOTIC APPROXIMATIONS

In this thesis, the term alpha value is used in a general sense, and refers to the probability of rejecting a true null hypothesis. The term P-value refers to the probability that a test statistic will exceed (or not exceed in the lower-tailed test) the computed value, when the hypothesis being tested is true.

For selected values, the exact cumulative distribution functions (C.D.F.) of the test statistics are compared with those obtained from normal based asymptotic approximations. The results of the comparisons are used as a basis for assessing the accuracy of the approximations. In those cases where more than one asymptotic approximation has been suggested in the literature, the accuracy of each approximation is compared over a range of desired C.D.F. values and sample sizes. From the results, the most consistently accurate approximation, and the sample size for which that approximation provides at least three decimal place accuracy is determined.

Once the accuracy comparisons were completed for a specific nonparametric test, microcomputer capabilities were considered. In some cases, generation of the exact distribution up to the desired sample size took too long or was not possible on the PC. When this occurred, the mainframe computer was used to generate the required distributions with the results stored in numerical matrices for quick recall by the nonparametric test programs.

IV. TESTS FOR LOCATION BASED ON SINGLE AND PAIRED-SAMPLE DATA

The tests assume that the data consists of a single set of independent observations X_i or paired observations (X_i, Y_i), $i=1,2,\ldots,N$, from a continuous distribution. For the single and paired-sample cases, the null hypotheses are concerned with the median of the X_i and the median of the differences $X_i - Y_i$, respectively. The tests considered are the Ordinary Sign Test and the Wilcoxon Signed-Rank Test.

A. ORDINARY SIGN TEST

The Sign test can be used to test various hypothesis about the population median (or the median of the population of differences). Confidence intervals for these parameters can also be constructed. As a final option, nonparametric confidence intervals for the quantiles of a continuous distribution are offered.

1. Computation of the Test Statistic

For single-sample data, the test statistic K is computed as the number of observations X_i greater than the hypothesized median M_0. For the paired-sample case, K is the number of differences $X_i - Y_i$ that exceed M_0. All observations X_i (or $X_i - Y_i$) that are equal to

M_0 are ignored and the sample size decreased accordingly. As long as the number of such ties is small relative to the size of the sample, the test results are not greatly affected. Gibbons [Ref. 2:pp. 108].

2. The Null and Asymptotic Distribution of K

The null distribution of K is binomial with p = .5. In Table 1, the exact values of the C.D.F are compared with the corresponding approximate values using a normal approximation with and without continuity correction.

TABLE 1. C.D.F. COMPARISONS FOR THE SIGN TEST

PROB[K ≤ k]; FOR SAMPLE SIZE EQUAL TO 24.

TEST STAT. VALUE	5	6	7	8	9	10	11
EXACT C.D.F.	.00331	.01133	.03196	.07579	.15373	.27063	.41941
ERROR; NORMAL	.00117	.00417	.01134	.02456	.04339	.06352	.07786
ERROR; NORM. W/CC	¯.00068	¯.00104	¯.00114	¯.00073	.00001	.00048	.00028

PROB[K ≤ k]; FOR SAMPLE SIZE EQUAL TO 25.

TEST STAT. VALUE	6	7	8	9	10	11	12
EXACT C.D.F.	.00732	.02164	.05388	.11476	.21218	.34502	.50000
ERROR: NORMAL	.00266	.00774	.01795	.03400	.05352	.07077	.07926
ERROR; NORM. W/CC	¯.00088	¯.00111	¯.00092	¯.00031	.00032	.00044	.00000

As can be seen, for sample sizes greater than 25, a normal approximation with continuity correction is accurate to at least three decimal places.

3. Hypothesis Testing

P-values are computed for three basic hypotheses comparing the median of the population or the median of the population of differences M with a hypothesized median M_0. P-values are taken from the cumulative distribution of the binomial for the following tests of hypothesis.

a. One-sided Tests

(1) H0: $M = M_0$ Versus H1: $M < M_0$. The P-value equals $\Pr[K \le k]$, where k is the computed value of the test statistic.

(2) H0: $M = M_0$ Versus H1: $M > M_0$. The P-value equals $\Pr[K \ge k]$.

b. Two-sided Test.

(1) H0: $M = M_0$ Versus H1: $M \neq M_0$. The P-value equals twice the smaller value of a(1) or a(2), but does not exceed the value one.

For sample sizes greater than 25, a normal approximation with continuity correction is used.

4. Confidence Interval Estimation

Confidence intervals for the population median are based on the ordered observations in the sample. For paired-sample data, confidence bounds are obtained from the ordered differences of the pairs of data. A $100(1-\alpha)\%$ confidence interval is determined in the following manner. Let k be the number such that $\Pr[K \leq k] \leq (\alpha/2)$. Then, the (k+1)th and (N-k)th order statistics constitute the end points of the confidence interval. Gibbons [Ref. 1: pp.104].

For computing confidence intervals when sample size N is greater than 25, a normal approximation with continuity correction is used.

Also included under this test is an option to generate nonparametric confidence intervals for any specified quantile given a sample size N from a continuous distribution. The end points of the intervals are sample order statistics.

B. WILCOXON SIGNED-RANK TEST

The signed-rank test requires the added assumption that the underlying distribution is symmetric. This test uses the ranks of the differences $X_i - M_0$ (or $X_i - Y_i - M_0$) together with the signs of these differences to determine the test statistic. Confidence intervals for the median can also be constructed.

1. Computation of the Test Statistic

For single-sample data, the test statistic W is computed as follows.

$$\text{Let } Z_i = \begin{cases} 1 & \text{if } X_i - M_0 > 0 \\ 0 & \text{if } X_i - M_0 \leq 0 \end{cases}$$

and let $r_i = \text{rank}(|X_i - M_0|)$. Then, $W = \sum_{i=1}^{N} Z_i r_i$.

For paired sample data, W is calculated in the same manner, except the differences to be ranked are the paired-differences minus the hypothesized median. Zero differences are ignored and the sample size is decreased accordingly. When ties occur between ranks, the average value of the ranks involved are assigned to the tied positions. It has been shown that a moderate number of ties and zero differences has little effect on the test results.[1]

2. The Null and Asymptotic Distribution of W

The exact null distribution of W is given by: $\Pr[W = w] = u_N(w)/2^N$, $w = 0,1,2,\ldots,N(N+1)/2$, where $u_N(w)$ is the number of ways to assign plus and minus signs to the first N integers such that the sum of the positive integers equals w. It can be shown (see Gibbons [Ref. 1:pp. 112]) that $u_N(w)$, for successive values of N, can be computed using the recursive relationship:

$$u_N(w) = u_{N-1}(w-N) + u_{N-1}(w)$$

---

[1] For more information on the effects that zeros and tied ranks have on the Wilcoxon Signed-Rank Test, see Pratt and Gibbons [Ref. 3].

Exact C.D.F. values were compared with those obtained using the following asymtotic approximations: student's T with (N-1) degrees of freedom (T), student's T with continuity correction (TC), normal (Z), normal with a continuity correction (ZC), the average of T and Z as suggested by Iman [Ref. 4], and the average of TC and ZC.

As can be seen in Table 2 below, the average of TC and ZC gives the most consistently accurate results with three decimal place accuracy when the sample size exceeds 9.

TABLE 2. C.D.F. COMPARISONS FOR THE WILCOXON SIGNED-RANK TEST

PROB[W ≤ w]; FOR SAMPLE SIZE EQUAL TO 9.

TEST STAT. VALUE	3	5	6	8	9	12	14
EXACT C.D.F.	.00977	.01953	.02734	.04883	.06445	.12500	.17969
ERROR; NORMAL	¯.00067	.00046	.00204	.00591	.00958	.01824	.02272
ERROR; NORM. W/CC	¯.00243	¯.00247	¯.00167	.00023	.00269	.00693	.00806
ERROR; T DIST	.00518	.00608	.00673	.00701	.00817	.00826	.00818
ERROR; T W/CC	.00355	.00276	.00233	.00012	¯.00010	¯.00437	¯.00725
ERROR; AVE T/Z	.00225	.00327	.00438	.00646	.00888	.01325	.01545
ERROR; AVE TC/ZC	.00056	.00014	.00033	.00017	.00129	.00128	.00041

TABLE 2. (Continued)

PROB[W ≤ w]; FOR SAMPLE SIZE EQUAL TO 10.

TEST STAT. VALUE	5	7	8	10	12	15	17
EXACT C.D.F.	.00977	.01355	.02441	.04199	.06543	.11621	.16113
ERROR; NORMAL	¯.00115	.00023	.00099	.00476	.00837	.01490	.01888
ERROR; NORM. W/CC	¯.00270	¯.00219	¯.00198	.00043	.00229	.00558	.00710
ERROR; T DIST	.00399	.00512	.00525	.00665	.00661	.00661	.00681
ERROR; T W/CC	.00249	.00243	.00182	.00151	¯.00052	¯.00376	¯.00571
ERROR; AVE T/Z	.00142	.00267	.00312	.00571	.00749	.01075	.01284
ERROR; AVE TC/ZC	¯.00010	.00012	¯.00008	.00097	.00089	.00091	.00070

3. Hypothesis Testing

P-values are computed for three basic hypotheses comparing the median of the population or the median of the population of differences M with a hypothesized median M_0 as shown below.

a. One-sided Tests

(1) $H_0: M = M_0$ Versus $H_1: M < M_0$. The P-value equals $\Pr[W \leq w]$, where w is the computed value of the test statistic W.

(2) $H_0: M = M_0$ Versus $H_1: M > M_0$. The P-value equals $\Pr[W \geq w]$.

b. Two-sided Test

(1) $H_0: M = M_0$ Versus $H_1: M \neq M_0$. The P-value equals twice the smaller value of a(1) or a(2), but not exceeding the value one.

For sample sizes greater than 9, an average of the normal and student's T approximations, each with continuity correction, is used. Computations of the P-value for each of the alternative hypotheses are:

a. $H1: M < M0$

Let $P_{ZC} = \Pr[\, Z \le (w + .5 - \mu_w)/\ \sigma_w]$ and

$$\text{let} \quad P_{TC} = \Pr\left[T_{(N-1)} \le \frac{|w - \mu_w| - .5}{\left[\dfrac{N\sigma_w^2\ [\,|w - \mu_w| - .5]^2}{N - 1}\right]^{.5}} \right]$$

where Z is standard normal, $T_{(N-1)}$ has a student's T distribution with (N-1) degrees of freedom, $\mu_w = N(N+1)/4$ and $\sigma_w^2 = (N(N+1)(2N+1)/24))$. Then, the P-value for the test is $(P_{ZC} + (1 - P_{TC}))/2$ if w is less than μ_w and $(P_{ZC} + P_{TC})/2$, otherwise. The above formulas are obtained from those given by Iman [Ref. 4] after inclusion of a continuity correction.

b. $H1: M > M0$

The P-value equals $((1 - P_{ZC}) + P_{TC})/2$ if w is less than μ_w and $((1 - P_{ZC}) + (1 - P_{TC}))/2$, otherwise. The computation of P_{ZC} and P_{TC} is similar to the above except the sign of the continuity correction is changed.

c. $H1: M \neq M0$

The P-value equals twice the smaller value of a or b above, but not exceeding the value one.

4. Confidence Interval Estimation

For single-sample data, the confidence interval for the population median is based on the ordered averages of all pairs of observations $(X_i + X_j)/2$ such that $i \leq j$. A $100(1-\alpha)\%$ confidence interval is determined in the following manner. Let w be the number such that $\Pr[W \leq w] \leq (\alpha/2)$. Then, the (w+1)th and (m-w)th order statistics, where $m = N(N+1)/2$ or the total number of paired-averages, constitute the end points of the confidence interval. A confidence interval for paired-sample data is computed in the same manner, except the end points are taken from the paired-averages of the differences $X_i - Y_i$. Gibbons [Ref. 1:pp. 114-118].

For computing confidence intervals when sample sizes are greater than 9, a normal approximation with continuity correction is used.

V. TESTS BASED ON TWO OR MORE SAMPLES

The tests assume that the data consists of independent random samples from two or more continuous distributions. The general null hypothesis is that the samples are drawn from identical populations. The Mann-Whitney and Kruskal-Wallis tests are considered.

A. MANN-WHITNEY TEST

The Mann-Whitney test is based on the distribution of the test statistic U, which can be used to compare the equality of the population medians or variances for two samples.[1] The Mann-whitney test with a modified ranking scheme can be used to test for equality of variances if the population means or medians are assumed to be equal (Conover [Ref. 5:pp. 229-230]). If the medians differ by a known amount, the data can be adjusted before applying the test. A confidence interval for the difference in the medians of the two populations can also be estimated.

1. Computation of the Test Statistic

For the comparison of population medians, the test statistic U is computed from the combined ordered arrangement of observations X_i and Y_j, $i = 1,2,\ldots,N$; $j = 1,2,\ldots,M$. Let $r_i = \mathrm{rank}(X_i)$ in the combined ordered sample and $R_X = \sum_{i=1}^{N} r_i$. Then,

---

[1] The test statistic U and the method used to compute it are taken from Gibbons [Ref. 1:pp. 140-141].

$$U = R_X - M(M+1)/2.$$

For testing the equality of variances, the computation of U is similar except for the method of assigning ranks to the ordered sample. This method ranks the smallest value 1, largest value 2, second largest value 3, second smallest value 4, and so on , by two's, until the middle of the combined ordered sample is reached. For either test, tied ranks for the combined sample are assigned the average value of the ranks involved. A moderate number of ties has little effect on the test results .

2. The Null and Asymptotic Distribution of U

The exact null distribution of U is determined using a recursion algorithm due to Harding [Ref. 6].

Exact C.D.F. values were compared with approximate values obtained from the following asymtotic distributions: student's T with (n-2) degrees of freedom where $n = N + M$, the total number of observations in both samples (T), student's T with continuity correction (TC), normal (Z), normal with continuity correction (ZC), the average of T and Z

(Iman [Ref. 7]), and the average of TC and ZC. The results for various sample sizes are given in Table 3.

TABLE 3. C.D.F. COMPARISONS FOR THE MANN-WHITNEY TEST

PROB[U ≤ u]; FOR SAMPLE SIZES N EQUAL TO 9 AND M EQUAL TO 9.

TEST STAT. VALUE	14	17	19	21	23	27	29
EXACT C.D.F.	.00933	.01999	.02515	.04696	.06796	.12904	.17005
ERROR; NORMAL	¯.00026	.00100	.00163	.00441	.00632	.01243	.01511
ERROR; NORM. W/CC	¯.00146	¯.00114	¯.00083	.00026	.00130	.00354	.00436
ERROR; T DIST	.00237	.00337	.00372	.00464	.00525	.00676	.00779
ERROR; T W/CC	.00119	.00108	.00095	.00007	¯.00073	¯.00259	¯.00328
ERROR; AVE T/Z	.00105	.00219	.00270	.00453	.00603	.00959	.01145
ERROR; AVE TC/ZC	¯.00014	¯.00003	.00004	.00017	.00026	.00048	.00054

PROB[U ≤ u]; FOR SAMPLE SIZES N EQUAL TO 7 AND M EQUAL TO 12.

TEST STAT. VALUE	14	17	19	21	24	27	30
EXACT C.D.F.	.00853	.01792	.02732	.04156	.07111	.11342	.17012
ERROR; NORMAL	¯.00045	.00062	.00187	.00359	.00701	.01097	.01487
ERROR; NORM. W/CC	¯.00152	¯.00128	¯.00079	¯.00003	.00154	.00322	.00458
ERROR; T DIST	.00199	.00292	.00356	.00422	.00527	.00643	.00796
ERROR; T W/CC	.00094	.00092	.00068	.00026	¯.00066	¯.00178	¯.00262
ERROR; AVE T/Z	.00077	.00177	.00272	.00391	.00614	.00870	.01142
ERROR; AVE TC/ZC	¯.00029	¯.00018	¯.00005	.00011	.00044	.00072	.00098

TABLE 3. (Continued)

PROB[U ≤ u]; FOR SAMPLE SIZES N EQUAL TO 5 AND M EQUAL TO 17.

TEST STAT. VALUE	13	16	18	20	23	27	30
EXACT C.D.F.	.00965	.01933	.02916	.04245	.07013	.12440	.17935
ERROR; NORMAL	¯.00077	.00039	.00170	.00349	.00639	.01210	.01564
ERROR; NORM. W/CC	¯.00190	¯.00150	¯.00087	.00007	.00133	.00444	.00578
ERROR; T DIST	.00130	.00222	.00299	.00390	.00544	.00774	.00955
ERROR; T W/CC	.00017	.00023	.00023	.00021	.00007	¯.00025	¯.00051
ERROR; AVE T/Z	.00027	.00130	.00235	.00369	.00616	.00992	.01260
ERROR; AVE TC/ZC	¯.00086	¯.00063	¯.00032	.00014	.00097	.00209	.00264

PROB[U ≤ u]; FOR SAMPLE SIZES N EQUAL TO 3 AND M EQUAL TO 27.

TEST STAT. VALUE	7	10	12	15	19	23	26
EXACT C.D.F.	.00764	.01650	.02512	.04236	.07734	.12660	.17433
ERROR; NORMAL	¯.00265	¯.00099	.00072	.00389	.00874	.01342	.01680
ERROR; NORM. W/CC	¯.00363	¯.00254	¯.00133	.00089	.00405	.00665	.00831
ERROR; T DIST	¯.00121	.00031	.00173	.00414	.00741	.01026	.01253
ERROR; T W/CC	¯.00219	¯.00129	¯.00042	.00097	.00250	.00328	.00390
ERROR; AVE T/Z	¯.00193	¯.00034	.00122	.00402	.00808	.01184	.01466
ERROR; AVE TC/ZC	¯.00291	¯.00192	¯.00087	.00093	.00327	.00496	.00610

As can be seen from the tables, the average of ZC and TC gives the most consistently accurate results. For sample sizes NxM > 80, nearly three decimal place accuracy is obtained in all cases.

3. Hypothesis Testing

P-values are computed for three basic hypotheses comparing the medians or variances of the two populations as shown below.

a. One-sided Tests

(1) $H0$: $M_X = M_Y$ Versus $H1$: $M_X < M_Y$ or $H0$: $V_X = V_Y$ Versus $H1$: $V_X > V_Y$. The P-value equals $Pr[U \leq u]$, where u is the observed value of the test statistic.

(2) $H0$: $M_X = M_Y$ Versus $H1$: $M_X > M_Y$ or $H0$: $V_X = V_Y$ Versus $H1$: $V_X < V_Y$. The P-value equals $Pr[U \geq u]$.

b. Two-sided Test

(1) $H0$: $M_X = M_Y$ Versus $H1$: $M_X \neq M_Y$ or $H0$: $V_X = V_Y$ Versus $H1$: $V_X \neq V_Y$. The P-value equals twice the smaller value of a(1) or a(2), but not exceeding the value one.

For sample sizes NxM greater than 80, the average of the normal and student's T approximations, each with continuity correction, is used. Computations of the P-value for each alternative hypothesis are:

a. H1: $M_X < M_Y$ or $V_X > V_Y$

Let $P_{ZC} = \Pr[\, Z \leq (u + .5 - \mu_u)/\ \sigma_u]$ and

$$\text{let } P_{TC} = \Pr\left[T_{(n-2)} \leq \frac{|u - \mu_u| - .5}{\left[\frac{(N+M-1)\sigma_u^2}{N+M-2} - \frac{[|u-\mu_u|-.5]^2}{N+M-2}\right]^{.5}} \right]$$

where Z is standard normal, $T_{(n-2)}$ has a student's T distribution with (n-2) degrees of freedom, $\mu_u = N x M/2$ and $\sigma_u^2 = (N(M)(N+M+1))/12$. Then the P-value for the test is $(P_{ZC} + (1 - P_{TC}))/2$ for u less than μ_u and $(P_{ZC} + P_{TC})/2$, otherwise. The above formulas are obtained from those given by Iman [Ref. 7] after inclusion of the continuity correction.

b. H1: $M_X > M_Y$ or $V_X < V_Y$

The P-value equals $((1 - P_{ZC}) + P_{TC})/2$ if u is less than μ_u and $((1 - P_{ZC}) + (1 - P_{TC}))/2$, otherwise. The computation of P_{ZC} and P_{TC} is similar to the above except the sign of the continuity correction is changed.

c. H1: $M_X \neq M_Y$ or $V_X \neq V_Y$

The P-value equals twice the smaller value of a or b above, but not exceeding the value one.

4. Confidence Interval Estimation

Confidence intervals for the difference in medians, $(M_Y - M_X)$, are based on the ordered arrangement of the differences $(Y_j - X_i)$, $j = 1,2,...,M$;

$i = 1,2,\ldots,N$ for all i and j. A $100(1-\alpha)\%$ confidence interval is determined in the following manner. Let u be the number such that $\Pr[U \leq u] \leq (\alpha/2)$. Then, the (u+1)th and (m-u)th order statistics, where $m = N \times M$ or the total number of possible differences, constitute the end points of the confidence interval.

For computing confidence intervals when sample sizes NxM are greater than 80, a normal approximation with continuity correction is used.

B. KRUSKAL-WALLIS TEST

The Kruskal-Wallis test is a nonparametric analog of the one-way classification analysis of variance test for equality of several population medians. Gibbons [Ref. 1:pp. 199].

1. Computation of the Test Statistic

Calculations of the test statistic H center around the ordered arrangement of the combined samples from which the sum of ranks for each sample is derived. Let X_{ij}, $j=1,2,\ldots,n_i$ and $i=1,2,\ldots,k$, be independent random samples from k populations. Let $r_{ij} = \mathrm{rank}(X_{ij})$, $R_i = \sum_{j=1}^{n_i} r_{ij}$, and $N = \sum_{i=1}^{k} n_i$. Then,

$$H = (12/(N(N+1))\left\{\sum_{i=1}^{k} (R_i^2/n_i)\right\} - 3(N+1)$$

If ties occur in the combined sample, they are resolved by assigning the average value of the ranks involved. A correction based on the number of observations tied at a given rank and the number of ranks involved, is included in the calculations. A complete description of the correction factor is given in Gibbons [Ref. 2:pp. 178-179].

2. The Null and Asymptotic distribution of H

The null distribution of H is generated by enumeration. Each possible permutation of ranks is listed for the combined sample, and the corresponding H value computed. The frequency distribution of H is the total number of occurrences of each distinct H value. The H values are arranged in increasing order while maintaining the frequency pairings. The null distribution is obtained by dividing the cumulative frequencies by $n_1!n_2!...n_k!/N!$.

Due to computer limitations, generation of the exact distribution of H was only possible for k = 3 populations with n = 4 observations in each, and 4 populations with 3 observations in each. Most of the distributions were generated on the mainframe computer and saved in matrices for quick recall by the Kruskal-Wallis test program.

Exact C.D.F. values were compared with the corresponding approximate values using the following distributions: chi-square with (k-1) degrees of freedom (C), F distribution with (k-1) and (N-k) degrees of freedom (F), and F with (k-1) and (N-k-1) degrees of freedom (F1). The chi-square distribution uses the Kruskal-Wallis H statistic, while the F and F1 distributions use a modified H statistic, H1 = ((N-k)H)/(k-1)((N-1)-H); see Iman and Davenport [Ref. 8]. As can be seen in Table 4, F1 gives the most consistently accurate estimates.

TABLE 4. C.D.F. COMPARISONS FOR THE KRUSKAL-WALLIS TEST

PROB[H ≥ h]; FOR A GROUP OF 3 SAMPLES CONSISTING OF 4, 4, AND 3 OBS.

TEST STAT. VALUE	7.1439	6.7121	6.1818	5.5985	5.0530	4.2121	3.5985
EXACT C.D.F.	.00970	.01905	.02961	.04866	.07810	.12918	.17784
ERROR; CHISQUARE	¯.01840	¯.01582	¯.01585	¯.01220	¯.00184	.00746	.01241
ERROR; F DIST	.00304	.00736	.00836	.01113	.01820	.01696	.00990
ERROR; F W/¯1 DF	.00084	.00426	.00403	.00544	.01138	.00895	.00172

PROB[H ≥ h]; FOR A GROUP OF 3 SAMPLES CONSISTING OF 4, 4, AND 4 OBS.

TEST STAT. VALUE	7.6538	6.9615	6.5000	5.6923	4.9615	4.2692	3.5769
EXACT C.D.F.	.00762	.01939	.02996	.04866	.08000	.12190	.17299
ERROR; CHISQUARE	¯.01416	¯.01139	¯.00882	¯.00941	¯.00368	.00361	.00577
ERROR; F DIST	.00290	.00839	.01204	.01100	.01272	.01225	.00263
ERROR; F W/¯1 DF	.00149	.00600	.00890	.00647	.00708	.00592	¯.00384

PROB[H ≥ h]; FOR A GROUP OF 4 SAMPLES CONSISTING OF 3, 3, 2, AND 2 OBS.

TEST STAT. VALUE	7.6364	7.1818	7.0000	6.5273	6.0182	5.3818	4.8727
EXACT C.D.F.	.01000	.01921	.02921	.04921	.07984	.12984	.17952
ERROR; CHISQUARE	¯.04416	¯.04712	¯.04269	¯.03939	¯.03089	¯.01604	¯.00183
ERROR; F DIST	.00284	.00260	.00730	.00880	.01094	.01123	.00902
ERROR; F W/¯1 DF	¯.00172	¯.00494	¯.00147	¯.00307	¯.00375	¯.00570	¯.00843

PROB[H ≥ h]; FOR A GROUP OF 4 SAMPLES CONSISTING OF 3, 3, 3, AND 2 OBS.

TEST STAT. VALUE	8.0152	7.6364	7.1515	6.7273	6.1970	5.4697	4.9697
EXACT C.D.F.	.00961	.01831	.02974	.04948	.07805	.12740	.17571
ERROR; CHISQUARE	¯.03609	¯.03584	¯.03748	¯.03164	¯.02436	¯.01306	.00168
ERROR; F DIST	.00215	.00481	.00441	.00920	.01185	.01036	.01214
ERROR; F W/¯1 DF	¯.00133	¯.00019	¯.00269	.00030	.00098	¯.00230	¯.00091

A final accuracy comparison between the C and F1 approximations was conducted by computer simulation for 5 populations with 8 observations each. Initially, 30,000 permutations of the 40 ranks were randomly generated (no tie ranks allowed) and the H statistic calculated for each permutation. Then the empirically determined percentiles H_p for selected values of p between .01 and .18 were compared with the approximations given by the C and F1 distributions. The results are shown in Table 5. It can be seen that the F1 approximation compares well with the simulated results, giving three decimal place accuracy, while the C approximation is less accurate.

TABLE 5. C.D.F. COMPARISONS FOR THE KRUSKAL-WALLIS TEST USING COMPUTER SIMULATION

PROB[H ≥ h]; BASED ON 10000 GENERATED H'S FOR 5 SAMPLES OF 8 OBS. EACH.

TEST STAT. VALUE	12.229	11.065	10.248	9.212	8.129	7.030	6.232
C.D.F. VALUE	.01000	.02000	.03000	.05000	.08000	.13000	.18000
ERROR; CHISQUARE	¯.00573	¯.00584	¯.00646	¯.00601	¯.00696	¯.00432	¯.00246
ERROR; F W/¯1 DF	.00081	.00231	.00259	.00350	.00161	.00109	¯.00103

PROB[H ≥ h]; BASED ON 20000 GENERATED H'S FOR 5 SAMPLES OF 8 OBS. EACH.

TEST STAT. VALUE	12.315	11.054	10.184	9.163	8.129	7.034	6.220
C.D.F. VALUE	.01000	.02000	.03000	.05000	.08000	.13000	.18000
ERROR; CHISQUARE	¯.00516	¯.00596	¯.00745	¯.00716	¯.00696	¯.00413	¯.00334
ERROR; F W/¯1 DF	.00126	.00221	.00166	.00235	.00161	.00130	¯.00199

PROB[H ≥ h]; BASED ON 30000 GENERATED H'S FOR 5 SAMPLES OF 8 OBS. EACH.

TEST STAT. VALUE	12.305	10.976	10.147	9.179	8.168	7.072	6.265
C.D.F. VALUE	.01000	.02000	.03000	.05000	.08000	.13000	.18000
ERROR; CHISQUARE	¯.00522	¯.00684	¯.00802	¯.00677	¯.00563	¯.00213	¯.00020
ERROR; F W/¯1 DF	.00121	.00143	.00111	.00273	.00301	.00344	.00142

3. Hypothesis Testing

P-values for the test H0: the population medians are all equal versus H1: at least two population medians are not equal, are computed as: $Pr[H \geq h]$, where h is the value of the observed test statistic.

For three or more populations with at least 4 observations in each, the F1 approximation is used.

VI. TESTS FOR ASSOCIATION IN PAIRED-SAMPLES

The tests described herein assume that the data consists of independent pairs of observations (X_i,Y_i) from a bivariate distribution. The general null hypothesis is that of no association between X and Y. Kendall's B and Spearman's R are considered.

A. KENDALL'S B

1. Computation of the Test Statistic

The test statistic is computed by comparing each observation (X_i,Y_i) with all other observations (X_j,Y_j) in the sample. If the changes in X and Y are of the same sign, $\mathrm{sgn}(X_j - X_i) = \mathrm{sgn}(Y_j - Y_i)$, the pair (X_i,Y_i) and (X_j,Y_j) is "concordant" and a +1 is scored. If the signs are different, the pair is "discordant" and a -1 is scored. Any ties between either the X's or the Y's scores a zero for that pair. The sum of all scores divided by the total number of distinguishable pairs, $(N(N-1))/2$, gives B. If zeros are scored, the denominator is reduced by a correction factor which is based on the number of observations tied at a given rank and the number of ranks involved in each of the X and Y samples. A complete description of the correction for ties is given in Gibbons [Ref. 2:pp.

289]. The value of B ranges between 1, indicating perfect concordance, and -1, for perfect discordance. Gibbons [Ref. 1:pp. 209-225].

2. The Null and Asymptotic Distribution of B

The null distribution of B is derived from the following recursive formula given in Gibbons [Ref. 1:pp.216].

$$u(N+1,P) = u(N,P) + u(N,P-1) + u(N,P-2) + \ldots + u(N,P-N)$$

where u(N,P) denotes the number of P concordant pairings of N ranks. This formula is used to generate the frequency with which the possible values of P occur. Division by N! results in the probability distribution of P. Since, $B = (4P/(N(N-1)))-1$, the null distribution of B is easily determined.

Exact C.D.F. values were compared with those obtained using a normal approximation, with and without a continuity correction factor ($CC = 6/N(N^2-1)$, proposed by Pittman [Ref. 11] for the Spearman's R test). The results for various sample sizes are provided in Table 6. As can be seen, for sample sizes greater than 13, a normal approximation with continuity correction provides three decimal place accuracy.

TABLE 6. C.D.F. COMPARISONS FOR KENDALL'S B

PROB[B ≥ b]; FOR SAMPLE SIZE EQUAL TO 13.

TEST STAT. VALUE	0.5128	0.4615	0.4103	0.3590	0.3333	0.2564	0.2308
EXACT C.D.F.	.00748	.01524	.02363	.04999	.06443	.12593	.15309
ERROR; NORMAL	.00014	.00121	.00318	.00620	.00808	.01473	.01703
ERROR; NORM. W/CC	-.00013	.00073	.00239	.00497	.00658	.01223	.01415

PROB[B ≥ b]; FOR SAMPLE SIZE EQUAL TO 14.

TEST STAT. VALUE	0.4725	0.4286	0.4066	0.3626	0.2967	0.2527	0.2088
EXACT C.D.F.	.00964	.01778	.02359	.03973	.07858	.11656	.16541
ERROR; NORMAL	.00035	.00140	.00218	.00431	.00889	.01256	.01627
ERROR; NORM. W/CC	.00008	.00095	.00162	.00345	.00742	.01057	.01371

3. Hypothesis Testing

P-values for tests of no association between X and Y are computed for three types of alternative hypotheses. Because the distribution of B is symmetric, all probabilities can be taken from the upper tail using the absolute value of b, the observed value of the test statistic. Linear interpolation is used when b lies between tabulated values. The P-values are computed as follows.

a. One-Sided Alternatives

The one-sided alternative tested depends on the sign of b. A positive b will automatically test for direct association or concordance, while a negative b will test for indirect association or discordance. The P-value equals $\Pr[B \geq |b|]$.

b. Two-Sided Alternative

The P-value equals twice the probability computed for the one-sided hypothesis.

For sample sizes greater than 12, a normal approximation with continuity correction is used. The approximate P-values are then:

$1 - \Pr[\, Z \leq ((|b| - CC) - \mu_b)/\, \sigma_b]$, where Z is standard normal, CC is the continuity correction, $\mu_b = 0$, and $\sigma_b^2 = (4N + 10)/9N(N-1)$, for the one-sided test and twice this P-value for the two-sided test.

B. SPEARMAN'S R

The Spearman's R Test requires the added assumption that the underlying bivariate distribution is continuous. The test measures the degree of correspondence between rankings, instead of the actual variate values, and can be used as a measure of association between X and Y. Gibbons [Ref. 1:pp. 226].

1. Computation of the Test Statistic

The test statistic R is computed in the following manner. Let $r_i = \text{rank}(X_i)$ and $s_i = \text{rank}(Y_i)$ and $D_i = r_i - s_i$. Then,

$$R = 1 - \frac{6 \sum_{i=1}^{N} D_i^2}{N(N^2 - 1)}$$

where N is the size of the sample. If ties occur in X or Y, they are resolved by assigning the average value of the ranks involved. A correction factor, based on the number of observations tied at a given rank and the number of ranks involved, is included in the calculations. A complete description of the correction factor is given in Gibbons [Ref. 2:pp. 279]. The value of R ranges between 1, indicating perfect direct association, and -1, for perfect indirect association. Gibbons [Ref. 1:pp. 226-235].

2. The Null and Asymptotic Distribution of R

The null distribution of R for a given sample size N is generated by enumeration. The method, as presented in Kendall [Ref. 9], involves generation of an N by N array of all possible squared differences between any two paired ranks of X and Y. All N! permutations of N ranks are used to index values from the array. The sum of these indexed values for each permutation gives rise to N! sum of squared differences which are then converted to the R statistic. The frequency distribution of R is the total number of occurrences of each distinct value of R divided by N!.

Due to mainframe computer memory limitations in the APL environment, generation of the distribution of R was limited to sample sizes of 7 or less.[1] Using tables, provided by Gibbons [Ref. 2:pp. 417-418] to supplement computer computations, a numerical matrix, called PMATSP, was created to store the cumulative distributions of R for sample sizes less than 11. This matrix allows for quick recall of cumulative probabilities by the Spearman's R Test program.

Exact C.D.F. values were compared with those obtained using a student's T approximation with (N-2) degrees of freedom (see Glasser and Winter [Ref. 10]), and a normal approximation. Both normal and T approximations were computed with and without a continutity correction factor, $CC = 6/N(N^2-1)$ (Pittman [Ref. 11]). From the results presented in Table 7, the most consistently accurate approximation is given by the T distribution with a correction.

3. Hypothesis Testing

P-values for tests of no association between X and Y can be computed for three types of alternative hypotheses. Because the distribution of R is symmetric, all probabilities are taken from the upper tail using

---

[1] The memory capacity of the mainframe computer in the APL environment is limited to 2.5 megabytes.

TABLE 7. C.D.F. COMPARISONS FOR SPEARMAN'S R

PROB[R ≥ r]; FOR SAMPLE SIZE EQUAL TO 9.

TEST STAT. VALUE	0.7333	0.7167	0.6667	0.6000	0.5333	0.4333	0.3500
EXACT C.D.F.	.00861	.01843	.02944	.04840	.07376	.12496	.17929
ERROR; NORMAL	¯.00475	¯.00290	¯.00023	.00356	.00805	.01479	.01319
ERROR: NORM. W/CC	¯.00558	¯.00413	¯.00185	.00128	.00498	.01029	.01236
ERROR; T DIST	.00235	.00352	.00451	.00459	.00415	.00298	.00138
ERROR; T W/CC	.00153	.00208	.00251	.00175	.00042	¯.00212	¯.00479

PROB[R ≥ r]; FOR SAMPLE SIZE EQUAL TO 10.

TEST STAT. VALUE	0.7455	0.6727	0.6364	0.5636	0.4909	0.4061	0.3333
EXACT C.D.F.	.00870	.01948	.02722	.04814	.07741	.12374	.17437
ERROR; NORMAL	¯.00396	¯.00230	¯.00091	.00271	.00700	.01216	.01572
ERROR; NORM. W/CC	¯.00457	¯.00327	¯.00210	.00095	.00451	.00867	.01128
ERROR; T DIST	.00204	.00296	.00326	.00328	.00258	.00159	.00107
ERROR; T W/CC	.00144	.00185	.00134	.00114	¯.00035	¯.00230	¯.00361

the absolute value of r, the observed value of the test statistic. The P-values are computed as follows.

a. One-Sided Alternatives

The one-sided alternative tested depends on the sign of r. A positive r will test for direct association, while negative r tests for indirect association. The P-value equals $\Pr[R \geq |r|]$.

b. Two-Sided Alternative

The P-value equals twice the probability computed for the one-sided hypothesis.

For sample sizes greater than 10, an approximation based on the student's T distribution with (N-2) degrees of freedom and continuity correction, is used. The P-values are:

$$1 - \Pr[\ T_{(N-2)} \leq ((|r| - CC) - \mu_r)/\ \sigma_r],$$

where $T_{(N-2)}$ denotes the T distribution with (N-2) degrees of freedom, CC is the continuity correction, $\mu_r = 0$, and $\sigma_r{}^2 = (1 - (|r| - CC)^2)/(N-2)$, for the one-sided test, and twice this P-value for the two-sided test. Gibbons [Ref. 1:pp. 218].

VII. NONPARAMETRIC SIMPLE LINEAR REGRESSION[1]

Nonparametric Linear Regression assumes that the data consists of independent pairs of observations from a bivariate distribution and that the regression of Y on X is linear. The program estimates linear regression parameters based on the data samples. It then allows the user to input X values to predict the Y values. Hypothesis testing and confidence interval estimation for the slope of the regression equation is offered. If the estimated slope lies outside the confidence interval, an alternate regression equation is offered with an opportunity to input X values to predict the corresponding Y values.

A. COMPUTATION OF THE ESTIMATED REGRESSION EQUATION

The least squares method is used to estimate A and B in the regression equation $Y_i = A + BX_i + e_i$ $(i=1,2,...N)$, where e_i (unobservable errors) are assumed to be independent and identically distributed. A and B are computed from the following equations:

---

[1] Except for program design considerations, the information and concepts provided in the section are paraphrased from Conover [Ref. 12:pp. 263-271].

$$B = \frac{N \sum_{i=1}^{N} X_i Y_i - \sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i}{N \sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2}$$

$$A = \frac{\sum_{i=1}^{N} Y_i - B \sum_{i=1}^{N} X_i}{N}$$

B. HYPOTHESIS TESTING

P-values for testing hypotheses about the slope of the regression equation are based on the Spearman's rank correlation coefficient R between the X_i and $U_i = Y_i - B_0 X_i$, where B_0 is the hypothesized slope. The appropriate one-sided test of hypothesis, H0: $B = B_0$ versus H1: $B < B_0$ or H1: $B > B_0$, is automatically chosen based on the sign of the computed test statistic r (positive r tests, H1: $B > B_0$; negative r tests, H1: $B < B_0$). The P-value is computed as: $\Pr[R \geq |r|]$. P-values for two-sided tests, H0: $B = B_0$ versus H1: $B \neq B_0$, are also presented.

For sample sizes N greater than 10, P-values are approximated using a T distribution with (N-2) degrees of freedom and continuity correction.

C. CONFIDENCE INTERVAL ESTIMATION

$100(1-\alpha)\%$ confidence bounds for the slope parameter B are determined as follows. The n possible slopes, $S_{ij} = (Y_i - Y_j)/(X_i - X_j)$, are computed for all pairs of data (X_i, Y_i) and (X_j, Y_j) such that $i < j$ and $X_i \neq X_j$ and rearranged in increasing order to give $S^{(1)} \leq S^{(2)} \leq \ldots \leq S^{(n)}$. Let w be the $(1-\alpha/2)$ percentile of the distribution of Kendall's statistic with sample size n.[1] Let d be the largest integer less than or equal to $(n-w)/2$ and u the smallest integer greater than or equal to $(n+w)/2 + 1$. Then $S^{(d)}$ and $S^{(u)}$ are the desired lower and upper confidence bounds, respectively.

For sample sizes larger than 13, a normal approximation with continuity correction is used to estimate the confidence intervals.

If the slope of the estimated regression equation does not lie within the computed confidence interval, the program automatically calculates a new regression equation where the slope is the median of the two-point slopes S_{ij} and the intercept is the difference of the medians of the X and Y samples, $M_Y - M_X$.[2]

---

[1] Kendall's statistic is defined here as $N_c - N_d$, where N_c is the number of concordant pairs of observations and N_d is the number of discordant pairs. Conover [REF. 12:pp. 256].

[2] This procedure is recommended by Conover [REF. 12:pp. 256].

VIII. AREAS FOR FURTHER WORK

To create a more versatile and powerful software package, the NONPAR workspace could be expanded to include some or all of the following nonparametric tests: tests for randomness based on runs, Chisquare and Kolmogorov-Smirnov(K-S) Goodness-of-fit tests, Chisquare and K-S general two sample distribution tests, Chisquare test for independence, and the Friedman test for association.

LIST OF REFERENCES

1. Gibbons, J. D., Nonparametric Statistical Inference, McGraw-Hill, Inc., 1971.

2. Gibbons, J. D., Nonparametric Methods for Quantitative Analysis, Holt, Rinehart, and Winston, 1976.

3. Pratt, J. W. and Gibbons, J. D., Concepts of Nonparametric Theory, pp. 160-176, Springer-Verlag, Inc., 1981.

4. Iman, R. L., "Use of a T-statistic as an Approximation to the Exact Distribution of the Wilcoxon Signed-ranks Test Statistic," Communications in Statistics, vol. 3, no. 8, pp. 795-806, 1974.

5. Conover, W. J., Practical Nonparametric statistics, John Wiley and Sons, Inc., 1971.

6. Harding, E. F., "An Efficient, Minimal-storage Procedure for Calculating the Mann-Whitney U, Generalized U and Similar Distributions," Applied Statistics, vol. 33, no. 1, pp. 1-6, 1984.

7. Iman, R. L., "An Approximation to the Exact Distribution of the Wilcoxon-Mann-Whitney Rank Sum Test Statistic, " Communications in Statistics -- Theoretical Methods, A5, no. 7, pp. 587-598, 1976.

8. Iman, R. L. and Davenport, J. M., "New Approximations to the Exact Distribution of the Kruskal-Wallis Test Statistic," Communications in Statistics -- Theoretical Methods, A5, no. 14, pp. 1335-1348, 1976.

9. Kendall, M. G., Kendall, S. F., and Smith, B. B., "The Distribution of the Spearman's Coefficient of Rank Correlation in a Universe in Which All Rankings Occur an Equal Number of Times," Biometrika, vol. 30, pp. 251-273, 1939.

10. Glasser, G. J. and Winter, R. F., "Critical Values of the Coefficient of Rank Correlation for Testing the Hypothesis of Independence," Biometrika, vol. 48, pp. 444-448, 1961.

11. Pitman, E. J. G., "Significance Tests Which May be Applied to Samples From any Populations. II. The Correlation Coefficient Test," Journal of the Royal Statistical Society, Supplement 4. pp. 225-232, 1937.

12. Conover, W. J., Practical Nonparametric Statistics, 2d ed., John Wiley and Sons, Inc., 1980.

APPENDIX A

DOCUMENTATION FOR THE MICROCOMPUTER WORKSPACE

1. General Information

This appendix describes the organization and operation of the IBM-PC (or compatible) version of the workspace. Appendix C continues from where this appendix leaves off, to walk the user through each test by working practical examples.

Before proceeding any further, the user should refer to section II (Workspace Design Issues) for general information about workspace requirements and assumptions regarding its use.

To get started, enter the APL environment in the usual manner and load the NONPAR workspace.

2. Workspace Menus

This workspace is designed around the use of menus. They guide the user through the selection process of choosing a nonparametric test and a test option. Three types of menus are used; the main menu, test menus, and help menus.

a. The Main Menu

Within moments of loading the NONPAR workspace, the main menu will appear. It is titled Nonparametric Statistical Tests. This menu presents general information about the workspace. Its primary purpose is to list the choices of nonparametric tests available and provide an option which allows the user to exit the main menu into APL to copy data into the workspace or return to DOS. Each test choice is listed with some information about the test's area of application. To make a selection from the menu, move the cursor (using the cursor keys) to highlight the desired choice, and press enter. As a reminder to the user, a footnote at the bottom of the screen describes the procedure for entering a choice. Once a test has been selected from the main menu, a sub-menu appropriate to the test appears. To exit from any menu back to the main menu, press the Escape key.

b. Test Menus

The title of the test menu is the name of the nonparametric test chosen. The text portion of the menu gives a general overview of the test, to include, the method used to compute the test statistic, and a description of the various options that may be exercised. The third section consists of the list of test options available. These options include returning to the main menu or choosing the help menu. Test menus may have options listed in single or multiple-paged formats. The comment in the final block of the menu lets the user know if a certain menu is multiple-paged or not. To make a selection from a multiple-paged menu, use the page-up or page-down key to locate the desired option. Proceed with the scroll keys to highlight the choice, and press enter. Once a test option is entered, the user is prompted to input the data required to run the test. When the option for more information is selected, the help menu is displayed.

c. Help Menus

The title of the help menu usually begins with the words "More Information About..." followed by the title of the nonparametric test. The text portion of the menu explains the test and its options in greater detail. No choices are offered in the menu. To return to the test menu, press any key.

APPENDIX B

DOCUMENTATION FOR THE MAINFRAME COMPUTER WORKSPACE

1. General Information

This appendix describes the organization and operation of the mainframe computer workspace. To load a copy of the NONPAR workspace from the APL library, enter the APL environment and type:)LOAD 9 NONPAR. Within a few moments the variables LIST and DESCRIBE are displayed on the screen. These variables provide a description of the workspace.

2. The NONPAR Workspace

The NONPAR workspace consists of seven programs which call several subprograms during their execution. The exact syntax for each test and its corresponding nonparametric test name is given in the following format:

SYNTAX: Nonparametric Test and Application.

a. SIGN: Ordinary Sign Test for Location in Single and Paired-sample Data.

b. WILCOX: Wilcoxon Signed-rank Test for Location in Single and Paired-sample Data.

c. MANNWHIT: Mann-Whitney Test for Equal Medians or Variances in Two Independent Samples.

d. KRUSKAL: Kruskal-Wallis Test for Equal Medians in K Independent Samples.

e. KENDALL: Kendall's B; Measure of Association for Paired-sample Data.

f. SPEARMAN: Spearman's R; Measure of Association Between Rankings of Paired Data.

g. NPSLR: Nonparametric Simple Linear Regression; Least Squares.

The list presented above can be displayed at any time by typing: LIST.

For each test program, there exists a HOW variable that gives a full description of the test and the various options that may be exercised. To display any of the HOW variables, just enter the test program's name with the suffix HOW appended (i.e. SIGNHOW).

A test is run by entering the program's name. The user is immediately prompted to input data. Enter numerical data separated by spaces or as a variable to which the numbers have been previously assigned. Several of the tests require a considerable amount of prompting before all the necessary data has been entered.

APPENDIX C

WORKSPACE FAMILIARIZATION THROUGH PRACTICAL EXAMPLES

1. General Information

This appendix applies to both the mainframe and microcomputer workspaces. Its purpose is to acquaint the user with the organization of the programs and the type of prompts to be expected.

Extensive error checking has been included in the programs to ensure that the data is of the proper form. Should a program become suspended, clear the state indicator by entering:)RESET, check over the data for errors, and restart the program. 330 kilobytes of computer memory are needed to load APL and the NONPAR workspace; to avoid filling up the remaining workspace area, the user should minimize data storage in the NONPAR workspace. To exit a program at any time, press the Control and Escape keys, simultaneously.

2. Practical Examples

a. Sign Test

(1) Description of Problem 1. A Sinclair mine is manufactured to have a median explosive weight of not less than 16 ounces. The explosive weights of 15 mines, randomly selected from the production line, were

recorded as follows: 16.2 15.7 15.9 15.8 15.9 16 16.1 15.8 15.9 16 16.1 15.7 15.8 15.9 15.8.

(a) Is the manufacturing process packing enough explosives in the mines?

(b) What range of values can be expected for the median of the explosive weights 90% of the time.

(2) Solution. To see if the manufacturing process is meeting the specifications, we test the hypothesis Ho: M = 16 versus H1: M < 16.

(3) Workspace Decision Process.

(a) Microcomputer: Choose the Sign Test from the main menu, and the option, Single Sample; Test H0: M = Mo versus H1: M < Mo, from the test menu. Skip to the Program Interaction section below.

(b) Mainframe: Enter SIGN at the keyboard and receive the prompt:

DID YOU ENTER THIS PROGRAM FOR THE SOLE PURPOSE OF GENERATING CONFIDENCE INTERVALS FOR A SPECIFIED SAMPLE SIZE AND QUANTILE? (Y/N).

Enter N (If Y is entered, the user will go directly to this last option of the test). The next prompt is:

THE NULL HYPOTHESIS STATES - THE POPULATION MEDIAN (M) IS EQUAL TO THE HYPOTHESIZED MEDIAN (Mo); H0: M = Mo. WHICH ALTERNATIVE DO YOU WISH

```
TO TEST?   ENTER:  1 FOR H1: M < Mo;  2 FOR H1: M > Mo; 3 FOR H1: M ≠ Mo.
```

Enter 1. The next prompt is:

```
ENTER:   1 FOR SINGLE-SAMPLE PROBLEM; 2 FOR PAIRED-SAMPLE PROBLEM.
```

Enter 1.

(4) <u>Program Interaction</u>. The prompt is:

```
ENTER  THE DATA (MORE THAN TWO OBSERVATIONS ARE REQUIRED).
```

Enter the data separated by spaces or as a variable to which the data has been previously assigned. The next prompt is:

```
ENTER THE HYPOTHESIZED MEDIAN.
```

Enter 16. The following is dislayed.

```
COMPUTATIONS ARE BASED ON A SAMPLE SIZE OF: 13.

THE TOTAL NUMBER OF POSTIVE SIGNS IS: 3.

THE  P-VALUE  FOR  H0:  M = 16  Versus  H1: M < 16 IS:  .0461.
```

Consider a significance level of .05. Since the P-value of .0461 is less than .05, we reject H0: M = 16 in favor of H1: M < 16 and conclude that the manufacturing process is not packing enough explosives in the Sinclair mine. The next prompt is:

```
WOULD  YOU  LIKE A CONFIDENCE INTERVAL  FOR THE MEDIAN? (Y/N).
```

Enter Y (If N is entered, the progam asks if confidence intervals for a quantile are desired). The next prompt is:

ENTER THE DESIRED CONFIDENCE COEFFICIENT; FOR EXAMPLE: ENTER 95, FOR A 95% CONFIDENCE INTERVAL.

Enter 90. The following is displayed.

A 90% CONFIDENCE INTERVAL FOR THE MEDIAN OF THE POPULATION IS: (15.8 $\leq$ MEDIAN $\leq$ 16).

The next prompt is:

WOULD YOU LIKE CONFIDENCE INTERVALS FOR A SPECIFIED QUANTILE? (Y/N).

To see the form of the results, we generate confidence intervals for the 30th quantile. Sample size is automaticly set at the number of data points entered eariler. Enter Y (If N is entered, the mainframe program ends; or, the Sign test menu reappears). The next prompt is:

ENTER DESIRED QUANTILE; FOR EXAMPLE: ENTER 20, FOR THE 20TH QUANTILE.

Enter 30. The following is displayed.

```
ORDER STATISTICS | COEFFICIENTS
    3        8   |    .823160
    2        9   |    .949490
    1       10   |    .991600
```

***** THIS TABLE GIVES CONFIDENCE COEFFICIENTS FOR VARIOUS INTERVALS WITH ORDER STATISTICS AS END POINTS FOR THE 30TH QUANTILE.

The mainframe program ends. The menu-driven microcomputer program pauses for input from the keyboard by prompting:

PRESS ENTER WHEN READY.

Press Enter and the Sign test menu reappears.

b. Wilcoxon Signed-rank Test

(1) Description of Problem 2. A special training program is being considered to replace the regular training that Radio Telephone Operators receive. In order to evaluate the effectiveness of the new training program, proficiency tests were given during the third week of regular training. Twenty-four trainees were chosen at random and grouped into twelve pairs based on proficiency test scores. One member of each pair received specialized training while the other member received regular training. Upon graduation, the proficiency tests were given again with the following results.

Specially Trained Group (X): 60 50 55 71 43 59 64 49 61 54 47 70

Regularly Trained Group (Y): 40 46 60 53 49 57 51 53 45 59 40 35

(a) Does the special training program ensure higher scores?

(b) By what range of values can the scores of the two groups be expected to differ 95% of the time?

(2) Solution. To test the hypothesis that the special training program raises profficiency scores, we test H0: M(X-Y) = 0 versus H1: M(X-Y) > 0.

(3) Workspace Decision Process

(a) Microcomputer: Choose the Wilcoxon Signed-rank Test from the main menu, and the option, Paired-sample; Test H0: M = Mo versus H1: M > Mo, from the test menu. Skip to the Program Interaction section below.

(b) Mainframe: Enter WILCOX at the keyboard and receive the prompts:

```
THE NULL HYPOTHESIS STATES - THE POPULATION MEDIAN (M) IS EQUAL TO THE HYPOTHESIZED MEDIAN (Mo); H0: M = Mo. WHICH ALTERNATIVE DO YOU WISH TO TEST? ENTER: 1 FOR H1: M < Mo; 2 FOR H1: M > Mo; 3 FOR H1: M ≠ Mo.
```

Enter 2. The next prompt is:

```
ENTER: 1 FOR SINGLE-SAMPLE PROBLEM; 2 FOR PAIRED-SAMPLE PROBLEM.
```

Enter 2.

(4) Program Interaction. The prompt is:

```
ENTER X DATA (MORE THAN TWO OBSERVATIONS ARE REQUIRED).
```

Enter the X data separated by spaces. The next prompt is:

ENTER Y DATA (NUMBER OF Y ENTRIES MUST EQUAL NUMBER OF X ENTRIES).

Enter the Y data. The next prompt is:

ENTER THE HYPOTHESIZED MEDIAN FOR THE DIFFERENCES OF THE PAIRED DATA.

Enter 0. The following is dislayed.

COMPUTATIONS ARE BASED ON A SAMPLE SIZE OF: 12.

THE TOTAL SUM OF POSITIVE RANKS IS: 60.5.

THE P-VALUE FOR COMPARING THE MEDIAN OF THE POPULATION OF DIFFERENCES TO THE HYPOTHESIZED MEDIAN, H0: M(X-Y) = 0 Versus H1: M(X-Y) > 0, IS: .0505.

Consider a significance level of .05. Since the P-value of .0505 is greater than .05, we do not reject the null hypothesis that the two training cources are equally effective. However, due to the closeness in values, the choice of rejecting or not rejecting the null hypothesis is strictly a judgement call. The next prompt is:

WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE MEDIAN? (Y/N).

Enter Y (If N is entered, the mainframe progam ends; or, the Wilcoxon test menu reappears). The next prompt is:

```
ENTER THE DESIRED CONFIDENCE COEFFICIENT;
FOR EXAMPLE: ENTER 95, FOR A 95% CONFIDENCE INTERVAL.
```

Enter 95. The following is displayed.

```
A 95% CONFIDENCE INTERVAL FOR THE MEDIAN OF
THE POPULATION OF DIFFERENCES IS:
( -1 ≤ MEDIAN(X-Y) ≤ 16.5 ).
```

The mainframe program ends. The menu-driven microcomputer program pauses for input from the keyboard by prompting:

```
PRESS ENTER WHEN READY.
```

Press Enter and the Wilcoxon test menu reappears.

c. Mann-Whitney Test for Equality of Medians

(1) Description of Problem 3. A group of Army and Navy officers were given the Defense Language Aptitude test. From the results, 14 Army and 17 Navy officers' scores were randomly selected. These scores are listed below.

Army (X): 35 30 55 51 28 25 16 63 60 44 20 42 47 38.

Navy (Y): 54 26 41 43 37 34 39 50 46 49 45 33 29 36 38 42 34.

(a) Is there sufficient evidence to claim that Navy officers score higher on this test than Army officers?

(b) By what range of values can the scores between the two groups be expected to differ 90% of the time.

(2) Solution. To see if Navy officers score higher on the exam, we test H0: $M_x = M_y$ versus H1: $M_x < M_y$.

(3) Workspace Decision Process.

(a) Microcomputer: Choose the Mann-Whitney Test from the main menu, and the option, Test H0: $M_x = M_y$ versus H1: $M_x < M_y$, from the test menu. Skip to the Program Interaction section below.

(b) Mainframe: Enter MANNWHIT at the keyboard and receive the prompts:

```
DO YOU WISH TO COMPARE THE MEDIANS OR VARIANCES OF THE POPULATIONS? ENTER: 1 TO COMPARE MEDIANS; 2 TO COMPARE VARIANCES.
```

Enter 1. The next prompt is:

```
THE NULL HYPOTHESIS STATES - THE MEDIANS OF X AND Y ARE EQUAL; Mx = My. WHICH ALTERNATIVE DO YOU WISH TO TEST? ENTER:
1 FOR H1: Mx < My; 2 FOR Mx > My; 3 FOR Mx ≠ My.
```

Enter 1.

(4) Program Interaction. The prompt is:

```
ENTER X DATA (MORE THAN ONE OBSERVATION IS REQUIRED).
```

Enter the X data separated by spaces. The next prompt is:

```
ENTER Y DATA.
```

Enter the Y data. The following is displayed.

```
THE SUM OF THE X RANKS IS: 224. THE U STATISTIC EQUALS: 119.

THE P-VALUE FOR H0: Mx =My versus H1: Mx < My IS: .5078.
```

We do not reject the hypothesis of equal population medians and conclude the median of all Army scores is equal to the Navy's. The next prompt is:

```
WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE SHIFT IN LOCATION (My - Mx)? (Y/N).
```

Enter Y (If N is entered, the mainframe program ends; or, the Mann-Whitney test menu reappears). The next prompt is:

```
ENTER THE DESIRED CONFIDENCE COEFFICIENT; FOR EXAMPLE: ENTER 95, FOR A 95% CONFIDENCE INTERVAL.
```

Enter 95. The following is displayed.

```
A 95% CONFIDENCE INTERVAL FOR THE SHIFT IN LOCATION BETWEEN POPULATIONS X AND Y IS:

( -10 ≤ My-Mx ≤ 10 ).
```

The mainframe program ends. The microcomputer program pauses for input from the keyboard by prompting:

PRESS ENTER WHEN READY.

Press Enter and the Mann-Whitney test menu reappears.

d. Mann-Whitney Test for Equality of Variances

(1) Description of Problem. Referring to problem 3 in section c(1). Is there sufficient evidence to claim that Army scores vary more than Navy scores?

(2) Solution. To see if Army scores vary more, we test H0: Vx = Vy versus H1: Vx > Vy.

(3) Workspace Decision Process.

(a) Microcomputer: Choose the Mann-Whitney Test from the main menu, and the option, Test H0: Vx = Vy versus H1: Vx > Vy, from the test menu and receive the prompt:

ENTER THE DIFFERENCE OF THE MEANS OR MEDIANS (Mx - My).

Because we believe the population medians to be approximately equal, We enter 0. Skip to the Program Interaction section below.

(b) Mainframe: Enter MANNWHIT at the keyboard and receive the prompts:

DO YOU WISH TO COMPARE THE MEDIANS OR VARIANCES OF THE POPULATIONS? ENTER: 1 TO COMPARE MEDIANS; 2 TO COMPARE VARIANCES.

Enter 2. The next prompt is:

THE TEST TO COMPARE VARIANCES, REQUIRES THE TWO POPULATION MEANS OR MEDIANS TO BE EQUAL. IF THEY DIFFER BY A KNOWN AMOUNT, THE DATA CAN BE ADJUSTED BEFORE APPLYING THE TEST. ENTER THE DIFFERENCE OF MEDIANS (Mx - My) OR 900 TO QUIT.

We enter 0. The next prompt is:

THE NULL HYPOTHESIS STATES - THE VARIANCES OF X AND Y ARE EQUAL; Vx = Vy. WHICH ALTERNATIVE DO YOU WISH TO TEST? ENTER:

1 FOR H1: Vx < Vy; 2 FOR Vx > Vy; 3 FOR Vx ≠ Vy.

Enter 2.

(4) Program Interaction. The prompt is:

ENTER X DATA (MORE THAN ONE OBSERVATION IS REQUIRED).

Enter the X data separated by spaces. The next prompt is:

ENTER Y DATA.

Enter the Y data. The following is displayed.

THE SUM OF THE X RANKS IS: 166. THE U STATISTIC EQUALS: 61.

THE P-VALUE FOR H0: Vx = Vy versus H1: Vx > Vy IS: .0112.

Consider a significance level of .05. Since a P-value of .0112 is less than .05, we reject the null hypothesis of equal variances in favor of

Vx > Vy and conclude that Army scores do vary more than Navy scores.

The mainframe program ends. The microcomputer program pauses for input from the keyboard by prompting:

PRESS ENTER WHEN READY.

Press Enter and the Mann-Whitney test menu reappears.

e. Kruskal-Wallis Test

(1) Description of Problem 4. During a recent Monster Mash involving four Navy SEAL Teams, one of the events consisted of the number of pushups a man could do in 2 minutes. Eight men were chosen randomly from each Team. The following scores were recorded.

SEAL 1: 90 96 102 85 65 77 88 70.

SEAL 2: 64 79 99 95 87 74 69 97.

SEAL 3: 101 66 93 89 71 60 76 98.

SEAL 4: 72 78 73 81 83 92 94 86.

Are the different Seal Teams considered to be equally fit?

(2) Solution. To see if the Seal Teams are equally fit, we test the hypothesis that all the population medians are equal.

(3) Workspace Decision Process.

(a) Microcomputer: Choose the Kruskal-Wallis Test from the main menu; and, once the test menu is displayed, press Enter.

(b) Mainframe: Enter KRUSKAL at the keyboard.

(4) Program Interaction. The prompt is:

ENTER THE NUMBER OF POPULATIONS TO BE COMPARED (MUST BE GREATER THAN TWO).

Enter 4. The next prompt is:

ENTER YOUR FIRST SAMPLE.

Enter the SEAL 1 data separated by spaces. The next prompt is:

ENTER YOUR NEXT SAMPLE.

Enter the SEAL 2 data. The next prompt is:

ENTER YOUR NEXT SAMPLE.

Enter the SEAL 3 data. The next prompt is:

ENTER YOUR LAST SAMPLE.

Enter the SEAL 4 data. The following is displayed.

THE H STATISTIC EQUALS: .1335.

THE P-VALUE FOR H0: THE POPULATION MEDIANS ARE EQUAL versus H1: AT LEAST TWO POPULATION MEDIANS ARE NOT EQUAL IS: .98893.

We do not reject the null hypothesis that the population medians are equal and conclude that the SEAL Teams are equally fit.

The mainframe program ends. The microcomputer program pauses for input from the keyboard by prompting:

PRESS ENTER WHEN READY.

Press Enter and the Kruskal-Wallis test menu reappears.

f. Kendall's B

(1) Description of Problem 5. In order to determine if cold weather affects target marksmanship, Naval Special Warfare recorded small arms marksmanship scores and corresponding air temperatures for a period of one year. 20 men were chosen at random, and their scores averaged for different air temperatures. The average score for each air temperature is shown below.

Air temperature (X): 50 55 20 50 65 55 30 52 40 60.

Average scores (Y): 210 200 165 165 260 215 175 191 180 235.

Can it be said that colder temperatures have an effect on marksmanship scores? Is that effect positive or negative?

(2) Solution. We test the null hypothesis that no association exists between cold temperatures and marksmanship.

(3) Workspace Decision Process.

(a) Microcomputer: Choose Kendall's B Test from the main menu; and, once the test menu is displayed, press Enter.

(b) Mainframe: Enter KENDALL at the keyboard.

(4) Program Interaction. The prompt is:

```
ENTER X DATA (MORE THAN TWO OBSERVATIONS
ARE REQUIRED).
```

Enter the X data separated by spaces. The next prompt is:

```
ENTER Y DATA (NUMBER OF Y ENTRIES MUST
EQUAL NUMBER OF X ENTRIES).
```

Enter the Y data. The following is displayed.

```
KENDALL'S B EQUALS: .7817.
THE P-VALUE FOR H0: NO ASSOCIATION EXISTS
versus: H1: DIRECT ASSOCIATION EXISTS IS: .00045.
THE P-VALUE FOR THE TWO-SIDED TEST OF
HYPOTHESIS IS: .0009.
```

Since the P-value for the one-sided test equals .00045, we reject the null hypothesis that no association exists between temperatures and marksmanship in favor of direct association. We conclude that colder temperatures tend to cause lower marksmanship scores.

The mainframe program ends. The microcomputer program pauses for input from the keyboard by prompting:

PRESS ENTER WHEN READY.

Press Enter and Kendall's B test menu reappears.

g. Spearman's R

(1) Description of Problem 6. When fitness reports are written, officers of the same grade are ranked against each other based upon their demonstrated level of performance. Last marking period, the Commanding and Executive Officers separately ranked 9 Ensigns as shown below.

	Ensigns								
	A	B	C	D	E	F	G	H	I
CO (X):	6	4	1	5	2	8	3	7	9
XO (Y):	5	6	3	4	1	9	7	2	8

Does any association exist between the two sets of rankings?

(2) Solution. We test the null hypothesis that no association exists.

(3) Workspace Decision Process.

(a) Microcomputer: Choose Spearman's R Test from the main menu; and, once the test menu is displayed, press Enter.

(b) Mainframe: Enter SPEARMAN at the keyboard.

(4) Program Interaction. The prompt is:

```
ENTER X DATA (MORE THAN TWO OBSERVATIONS ARE REQUIRED).
```

Enter the X data separated by spaces. The next prompt is:

```
ENTER Y DATA (NUMBER OF Y ENTRIES MUST EQUAL NUMBER OF X ENTRIES).
```

Enter the Y data. The following is displayed.

```
SPEARMAN'S R EQUALS: .5500.
THE P-VALUE FOR H0: NO ASSOCIATION EXISTS versus: H1: DIRECT ASSOCIATION EXISTS IS: .0664.
THE P-VALUE FOR THE TWO-SIDED TEST OF HYPOTHESIS IS: .1328.
```

Consider a significance level of .05. Since a P-value of .0664 exceeds .05, we do not reject the null hypothesis that no correspondence exists between the two sets of rankings.

The mainframe program ends. The microcomputer program pauses for input from the keyboard by prompting:

PRESS ENTER WHEN READY.

Press Enter and the Spearman's R test menu reappears.

h. Nonparametric Simple Linear Regression; Least Squares

(1) Description of Problem 7. Battery-powered Swimmer Proplusion Units are sometimes used to aide swimmers during long underwater swims. Recent tests have shown that a nearly linear relationship exists between water temperature and battery life for these units. The following 17 data points were randomly selected from the test results.

Water temperature (X)	Battery life (Y)
70	3
65	2.75
50	1.8
40	1.2
60	2.4
55	1.9
52	1.75
50	1.7
43	1.6
40	1.1
72	2.75
55	2
48	1.5
35	.9
70	3.3
68	3
57	2.3

(a) Find the fitted regression equation.

(b) For the following water temperatures, predict the battery life of the units: 61 52 46 36.

(c) Can we determine with any certainty if the slope of the regression line equals .05.

(d) What range of values could be used as the slope of the estimated equation line 90% of the time?

(2) Solution. To determine the estimated regression equation, we use nonparametric linear regression.

(3) Workspace decision process.

(a) Microcomputer: Choose Nonparametric Simple Linear Regression from the main menu; and, once the test menu is displayed, press Enter.

(b) Mainframe: Enter NPSLR at the keyboard.

(4) Program Interaction. The prompt is:

ENTER X DATA (MORE THAN TWO OBSERVATIONS ARE REQUIRED).

Enter the X data separated by spaces. The next prompt is:

ENTER Y DATA (NUMBER OF Y ENTRIES MUST EQUAL NUMBER OF X ENTRIES).

Enter the Y data. The following is displayed.

THE LEAST SQUARES ESTIMATED REGRESSION EQUATION IS:

Y = -1.263 + .060668X.

The next prompt is:

DO YOU WISH TO ENTER SOME X VALUES TO GET THE PREDICTED Y'S? (Y/N).

Enter Y (If N is entered, the program skips to hypothesis testing for the slope).

ENTER X VALUES.

Enter 61 52 46 36. The next prompt is:

THE PREDICTED Y VALUES ARE: 2.44 1.89 1.53 .92.

WOULD YOU LIKE TO RUN SOME MORE X VALUES?

Enter N. The next prompt is:

WOULD YOU LIKE TO TEST HYPOTHESIS ON B, THE SLOPE OF THE EQUATION? (Y/N).

Enter Y (If N is entered, the program skips to confidence interval estimation). The next prompt is:

ENTER THE HYPOTHESIZED SLOPE.

Enter .05. The following is displayed.

SPEARMAN'S R EQUALS: .5756.

THE P-VALUE FOR H0: B = .05 versus H1: B > .05 IS: .0079.

THE P-VALUE FOR THE TWO-SIDED TEST OF HYPOTHESIS IS: .0158.

Consider a significance level of .05. Since a P-value of.0079 is less than .05, we reject the null hypothesis that B = .05 in favor of B > .05, and conclude the slope of the regression line is greater than .05. The next prompt is:

WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE SLOPE? (Y/N).

Enter Y (If N is entered, the mainframe program ends; or, the Nonparametric Regression test menu reappears). The next prompt is:

ENTER THE DESIRED CONFIDENCE COEFFICIENT; FOR EXAMPLE: ENTER 95, FOR A 95% CONFIDENCE INTERVAL.

Enter 90. The following is displayed.

A 90% CONFIDENCE INTERVAL FOR B, THE SLOPE OF THE ESTIMATED REGRESSION LINE, IS:

$(.05333 \leq B \leq .07)$.

If the estimated slope, does not lie within the confidence interval, the following would be displayed.

THE LEAST SQUARES ESTIMATOR OF B LIES OUTSIDE THE CONFIDENCE INTERVAL. DISCARD THE LEAST SQUARES EQUATION AND USE:

$Y = -1.4458 + .060833X$.

THIS EQUATION IS BASED ON THE MEDIANS OF THE X AND Y DATA, AND THE MEDIAN OF THE TWO-POINT SLOPES CALCULATED FOR THE CONFIDENCE INTERVAL ON B.

The next prompt is:

DO YOU WISH TO ENTER SOME X VALUES TO GET THE PREDICTED Y'S FROM THE NEW EQUATION? (Y/N).

To compare results, let us input the temperatures in the new equation. Enter Y (If N is entered, the mainframe program ends; or, the Nonparametric Regression test menu reappears). The next prompt is:

ENTER X VALUES.

Enter 61 52 46 36. The following is displayed.

THE PREDICTED Y VALUES ARE: 2.265 1.72 1.35 .74.

The next prompt is:

WOULD YOU LIKE TO RUN SOME MORE X VALUES?

Enter N. The next prompt is:

WOULD YOU LIKE TO TEST HYPOTHESIS ON B, THE SLOPE OF THE EQUATION? (Y/N).

To compare results once again, we enter Y (If N is entered, the mainframe program ends; or, the Nonparametric Regression test menu reappears). The next prompt is:

ENTER THE HYPOTHESIZED SLOPE.

Enter .05. The following is displayed.

SPEARMAN'S R EQUALS: .8287.

THE P-VALUE FOR H0: B = .05 versus H1: B > .05 IS: .0000.

THE P-VALUE FOR THE TWO-SIDED TEST OF HYPOTHESIS IS: .0000.

The mainframe program ends. The microcomputer program pauses for input from the keyboard by prompting:

PRESS ENTER WHEN READY.

Press Enter and the Nonparametric regression test menu reappears.

APPENDIX D

MAIN PROGRAM LISTINGS FOR MICROCOMPUTER WORKSPACE

```
     ∇ KEN;A;AA;B;BX;BY;C;CX;CY;D;DD;DX;DY;DXY;S;POS;NEG;XX;YY;N;DEN;NN;NUM;P;
       PVAL;SU;SV;T;U;V;AT;Z;X;Y;WW;CHA;E;PV;Q;R
⍝   THIS FUNCTION COMPUTES THE KENDALL B STATISTIC WHICH IS A MEASURE
⍝ OF ASSOCIATION BETWEEN SAMPLES. P-VALUES ARE GIVEN FOR TESTING ONE
⍝ AND TWO-SIDED HYPOTHESIS FOR NO ASSOCIATION VERSUS ASSOCIATION.
⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: TIES, TIESK, KENDALP,
⍝ INTERP, INPUT AND NORMCDF.

⍝          DISPLAY TEST MENU AND INPUT DATA
N1:E←MENU KENOBJ
 →(E=1)/B1
 MENU MAINOBJ
 →0
B1:R←INPUT 2
 Q←1↑R
 X←1↓(Q+1)↑R
 Y←(Q+1)↓R
⍝             ORDER Y IN INCREASING ORDER OF X
 A←Y[⍋X]
⍝            ORDER X IN INCREASING ORDER
 B←X[⍋X]
⍝              COMPUTE CURRENT RANKING OF Y
 C←⍋⍋A
⍝              NOW ORDER Y RANKS IN INCREASING ORDER
 D←A[⍋A]
⍝      IF TIES EXIST IN EITHER X OR Y RANKED VECTOR USE MID-RANK METHOD
 DD←1 TIES D
 XX←1 TIES B
⍝        FIND ORIGINAL RANKING OF Y WITH TIES RESOLVED
 YY←DD[C]
 N←ρX
⍝ COMPUTE NUMBER OF DISTINGUISHABLE PAIRS
 NN←(N×(N-1))÷2
 S←ρ0
 AA←0
⍝ POSITIVE ONES COME FROM A RUNS UP CONDITION; NEGATIVE 1 FROM RUNS DOWN
⍝ ZERO IS SCORED FOR TIES. MULTIPLY THE RESULTS FOR EVERY ELEMENT AND SUM
L1:AA←AA+1
 BX←(XX[AA]>(AA↓XX))
 CX←(XX[AA]<(AA↓XX))×(¯1)
 DX←BX+CX
 BY←(YY[AA]>(AA↓YY))
 CY←(YY[AA]<(AA↓YY))×(¯1)
 DY←BY+CY
 DXY←DX×DY
 POS←(DXY>0)
 NEG←(DXY<0)×(¯1)
 S←S,POS,NEG
 →(AA<(N-1))/L1
⍝            SUM FINAL VECTOR TO DETERMINE S
 S←+/S
⍝    OBTAIN THE NUMBER OF TIES IN EACH VECTOR USING THE TIESK FUNCTION
 U←TIESK B
 V←TIESK D
 SU←+/(2!U)
 SV←+/(2!V)
⍝ CALCULATE THE B STATISTIC INCLUDING THE CORRECTION FOR TIES
 T←S÷((NN-SU)×(NN-SV))*0.5
 AT←|T
 →(N>13)/NORM
⍝         CALL KENDALP TO CALCULATE THE RIGHT TAIL OF THE CDF OF B
 P←KENDALP N
⍝ CALL INTERP TO CALCULATE P-VALUE BY INTERPOLATION
 PVAL←AT INTERP P
 →(PVAL=¯1)/L3
 PVAL←0.5
 →L3
⍝ CALCULATE P VALUE USING NORMAL APPROX.
NORM:NUM←(3×AT)×((2×NN)*0.5)
 DEN←(2×((2×N)+5))*0.5
 Z←NUM÷DEN
 PVAL←1-(NORMCDF Z)
⍝ IF B IS POSITIVE PRINT OUT DIRECT ASSOCIATION.
L3:→(T>0)/L5
 CHA←'INDIRECT'
 →L7
L5:CHA←'DIRECT'
L7:PV←2×PVAL
 →(PV≤1)/L8
 PV←1
L8:'KENDALL''S B EQUALS: ',(4⍕T),⎕TCNL
 'THE P-VALUE FOR H0: NO ASSOCIATION EXISTS VERSUS'
 '                H1: ',(⍕CHA),' ASSOCIATION EXISTS IS: ',(4⍕PVAL),⎕TCNL
 'THE P-VALUE FOR THE TWO-SIDED TEST OF HYPOTHESIS IS: ',(4⍕PV),⎕TCNL
 'PRESS ENTER WHEN READY.'
 WW←⎕
 →N1
 ∇
```

```
 ∇ KRWL;NUM;DENOM;A;C;H;D;K;AA;BB;DD;E;F;N;OF;P;PVAL;R;SOFR;SR;TSOR;CHA;B
⍝     THIS FUNCTION COMPUTES THE KUSKAL-WALLIS TEST STATISTIC H WHICH IS
⍝ A MEASURE OF THE EQUALITY OF K INDEPENDENT SAMPLES.
⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: TIES, TIESK, INDEXPLS,
⍝ FDISTN, INTERP AND THE VARIABLES PMATKW20, PMATKW31, PMATKW33, PMATKW34
⍝ PMATKW41, PMATKW42, AND PMATKW43.

⍝   MENU CHOICES AND ROUTE TO PROPER STATEMET FOR ACTION.
N1:B←MENU KRWLQBJ
 →(B=1)/B1
 MENU MAINQBJ
 →0
 ⎕PP←4
B1:'ENTER THE NUMBER OF POPULATIONS TO BE COMPARED (MUST BE GREATER THAN T
 WO).'
 K←⎕
 →((K<3)∨((1|K)≠0))/E1
 →((ρK)>1)/E1
⍝                INITIALIZE VECTORS E AND F AND VARIABLE C
 E←F←SOFR←ρ0
 C←0
⍝THIS LOOP FACILITATES ENTERING THE SAMPLE VECTORS AND STORING THEM
 CHA←'FIRST'
L1:C←C+1
 'ENTER YOUR ',(⍕CHA),' SAMPLE.'
 D←⎕
 →((ρρD)=0)/NEXT
 D←1ρD
⍝   CONCATENATE SAMPLES AS THEY ARE ENTERED AND STORE THEM IN VECTOR E
NEXT:E←E,D
⍝         RECORD THE LENGTHS OF THE SAMPLES AS THEY ARE ENTERED
 F←F,ρD
 CHA←'NEXT'
 →(C<(K-1))/L1
 CHA←'LAST'
 →(C<K)/L1
⍝        RECORD SIZE OF ALL SAMPLES WHEN COMBINED
 N←+/F
⍝ ORDER SAMPLE SIZES LARGEST TO SMALLEST
 OF←F[⍒F]
⍝          ORDER COMBINED SAMPLE VECTOR TO BE USED BY TIES FUNCTION
 D←E[⍋E]
⍝   CALL INDEXPLS TO INCREMENT INDEXES WHEN TIES OCCUR WITHIN ONE SAMPLE
 AA←F INDEXPLS E
⍝            CALL TIES TO BREAK TIES BY MIDRANK METHOD
 BB←1 TIES D
 C←0
⍝             THIS LOOP CALCULATES THE H STATISTIC
L2:C←C+1
⍝             SUM OF RANKS FOR EACH SAMPLE IS CALCULATED
 SR←+/BB[(F[C]↑(AA[C;]))]
⍝  CALCULATE SUM OF RANKS SQUARED DIVIDED BY THE INDIVIDUAL SAMPLE SIZE
 SR←(SR*2)÷F[C]
⍝               STORE EACH CALCULATION
 SOFR←SOFR,SR
 →(C<K)/L2
⍝               SUM ACROSS ALL SAMPLES
 TSOR←+/SOFR
⍝              CALCULATE FINAL H STATISTIC
 H←(TSOR×(12÷(N×(N+1))))-(3×(N+1))
⍝  RECALCULATE H WITH CORRECTION FOR TIES
 A←TIESK E
 NUM←(+/(A*3))-(+/A)
 DENOM←N×((N*2)-1)
 H←H÷(1-(NUM÷DENOM))
⍝  SYSTEM OF LOGICAL STATEMENTS ENSURE PROPER PROB. IS ACCESSED
 →(OF[1]<2)/OUTPUT
 →(K>4)/FAPPROX
 →(K=3)/IF
 →((∧/(OF= 3 3 3 3))∨(OF[1]>3))/FAPPROX
 →(∧/(OF= 2 1 1 1))/OUTPUT
 →((K=4)∧(OF[1]=2))/P42
 →((∧/(OF= 3 ¯1 1 1))∨(∧/(OF= 3 2 1 1))∨(∧/(OF= 3 3 1 1))∨(∧/(OF= 3 3 2 1))
  )/P41
 →(K=4)/P43
IF:→(OF[1]>4)/FAPPROX
 →((∧/(OF= 2 1 1))∨(∧/(OF= 3 1 1)))/OUTPUT
 →((OF[1]=4)∧(OF[3]=1))/P31
 →((∧/(OF= 3 2 1))∨(∧/(OF= 3 3 1))∨(∧/(OF= 3 3 2))∨(∧/(OF= 3 3 3)))/P33
 →((∧/(OF= 3 2 2))∨(OF[1]=4))/P34
⍝             CALL APPROPRIATE VARIABLE ACCESS CDF
P23:P←PMATKW20[(N-4);;]
 →PM
P31:P←PMATKW31[(N-5);;]
 →PM
P33:P←PMATKW33[(N-5);;]
 →PM
P34:P←PMATKW34[(N-6);;]
 →PM
P41:P←PMATKW41[(N-5);;]
 →PM
P42:P←PMATKW42[(N-5);;]
 →PM
P43:P←PMATKW43[(N-7);;]
⍝    CALL INTERP TO CALCULATE P-VALUE BY INTERPOLATION
PM:PVAL←H INTERP P
 →(PVAL=¯1)/OUTPUT
 →L5
⍝  CALCULATE P-VALUE USING THE F DIST W/ONE LESS D.F.IN DENOM APPROX.
FAPPROX:F←((N-K)×H)÷((K-1)×((N-1)-H))
 PVAL←1-(((K-1),((N-K)-1)) FDISTN F)
 →L5
OUTPUT:PVAL←'GREATER THAN .25'
L5:'THE H STATISTIC EQUALS: ',(4⍕H),⎕TCNL
 'THE P-VALUE FOR H0: THE POPULATION MEDIANS ARE EQUAL VERSUS ',⎕TCNL
 '        H1: AT LEAST TWO POPULATION MEDIANS ARE NOT EQUAL IS: ',(4⍕PVAL),
 ⎕TCNL
 'PRESS ENTER WHEN READY.'
 AA←⍞
 →N1
E1:'ERROR: ENTER A SINGLE INTEGER VALUE GREATER THAN 2; TRY AGAIN.',⎕TCNL
 →B1
∇


∇ MANW;N;M;PV2;A;B;C;G;MM;NN;RX;U;NM1;P;NU;PVAL;NM;NUMZ;NUMZ1;DEN;DENC;DE
  NC1;TC;TC1;NUM;Z;Z1;ALPHA;CDF;INDEX;IPX;CI;UALPHA;BB;CC;U1;U2;PV;NN1;NN
  2;PVI;DIFF;AX;AY;AA;GG;AX1;PVM;PV3;D;Q;R
⍝   THIS FUNCTION USES THE SUM OF RANKS PROCEDURE TO CALCULATE THE MANN-
⍝ WHITNEY U STATISTIC WHICH IS USED IN COMPUTING THE P-VALUE FOR THE TEST
⍝ OF LOCATION AND SCALE.  THE  C.I. FOR (MY-MX), THE SHIFT IN LOCATION IS
⍝ ALSO COMPUTED. SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: TIES, TIES2
⍝ INDEXPLS, VARMW, MANWP, INPUT, CONFMW, NORMCDF, AND NORMPTH.
 →N3
N1:MENU MANWHELP
⍝ MENU CHOICES AND ROUTE TO PROPER STATEMENTS FOR ACTION.
N3:D←CHOICEM PAGEDMENU MANWOBJ
 DIFF←0
 →(D=2,3,4,5)/B1
 →(D=6,7,8)/B3
 →(D=1)/N1
 MENU MAINOBJ
 →0
B3:'ENTER THE DIFFERENCE OF THE MEANS OR MEDIANS (MX - MY).'
 DIFF←⎕
 →((⍴DIFF)>1)/E3
⍝                  ENTER DATA VECTORS
B1:'ENTER X DATA (MORE THAN ONE OBSERVATION IS REQUIRED).'
 N←⎕
 →((⍴⍴N)=0)/E1
 'ENTER Y DATA.'
 M←⎕
⍝ IF CALCULATIONS INVOLVE VARIANCES ADJUST X BY THE DIFFERENCE IN MEANS
 N←N-DIFF
⍝         CONCATENATE X AND Y SAMPLE VECTORS
 A←N,M
⍝         DETERMINE SIZE OF X AND Y VECTORS AND ASSIGN TO NN AND MM
 NN←⍴N
 MM←⍴M
⍝         COMPUTE SIZE LIMIT OF LEFT TAIL OF NULL DISTRIBUTION
 NM←(NN×MM)÷2
 NM1←⌊NM
⍝             ORDER A AND ASSIGN TO B
 B←A[⍋A]
 C←(NN,MM) INDEXPLS A
⍝     CALL TIES FUNCTION TO BREAK TIES USING MIDRANK METHOD
 G←1 TIES B
⍝ IF FALSE CALCULATE TEST FOR VARIANCES
 →(D=2,3,4,5)/B5
⍝     CALL VARMW TO GENERATE RANKS REQUIRED FOR VARIANCE TEST
 GG←VARMW(NN+MM)
⍝ CALL TIES TO RECORD TIES IN THE DATA AND BREAK TIES IN GG
```

```
 G←GG TIES B
⍝              CALCULATE SUM OF X RANKS
B5:RX←+/(G[(C[1;(⍳NN)])])
⍝             CONVERT TO MANNWHIT U STATISTIC
 U←RX-((NN×(NN+1))÷2)
 U1←U
⍝      IF SIZE OF X TIMES SIZE OF Y > 80; GO TO NORMAL APPROX
 →((NM×2)>80)/L2
 NN1←⌊/NN,MM
 NN2←⌈/NN,MM
⍝MANWP FUNCTION CALCULATES LEFT TAIL CUMULATIVE PROBS. OF U STATISTIC
 P←NN2 MANWP NN1
 →(D=5)/L10
⍝      LOGICAL STATEMENT ENSURES ONLY LEFT SIDE OF NULL DIST IS USED
 →(U≤NM1)/L3
⍝       CONVERT U STAT WHEN GREATER THAN LEFT TAIL VALUES
 U←(NN×MM)-U
⍝    IF U1 IS A FRACTIONAL, INTERPOLATE P VALUE
L3:→((1|U)=0)/NON
 U2←⌊U
 →(U2>0)/P1
 PV←1-((P[U2+1])÷2)
 →P3
P1:PV←1-((P[U2]+P[U2+1])÷2)
P3:PVI←(P[U2+1]+P[U2+2])÷2
 →CHECK
NON:→(U>0)/GO
 PV←1
 →P2
GO:PV←1-P[U]
P2:PVI←P[(U+1)]
CHECK:→(U1≤NM1)/L4
 PV2←PV
 PV←PVI
 PVI←PV2
 →L4
⍝ IF CONFIDENCE INTERVAL DESIRED GO TO L10
L2:→(D=5)/L10
⍝ COMPUTE THE  NORMAL APPROXIMATION W/CORRECTION FACTOR
 NUMZ←(U+0.5)-NM
 NUMZ1←(U-0.5)-NM
 DEN←((MM×NN×(MM+NN+1))÷12)*0.5
 Z←NUMZ÷DEN
 Z1←NUMZ1÷DEN
 NUM←|(U-NM)
 DENC←((((NN+MM-1)×(DEN*2))÷(NN+MM-2))-(((NUM-0.5)*2)÷(NN+MM-2)))*0.5
 DENC1←((((NN+MM-1)×(DEN*2))÷(NN+MM-2))-(((NUM+0.5)*2)÷(NN+MM-2)))*0.5
 TC←(NUM-0.5)÷DENC
 TC1←(NUM+0.5)÷DENC1
 →(U≤NM)/SECOND
 PVI←((NORMCDF Z)+((NN+MM-2) TDISTN TC))÷2
 PV←((1-(NORMCDF Z1))+(1-((NN+MM-2) TDISTN TC1)))÷2
 →L4
SECOND:PVI←((NORMCDF Z)+(1-((NN+MM-2) TDISTN TC)))÷2
 PV←((1-(NORMCDF Z1))+((NN+MM-2) TDISTN TC1))÷2
L4:PV3←2×(⌊/(PV,PVI))
 →(PV3≤1)/N5
 PV3←1
N5:PVM←(3,1)⍴(PVI,PV,PV3)
 'THE SUM OF THE X RANKS IS: ',(⍕RX),'.  THE U STATISTIC EQUALS: ',(⍕U1),
 ⎕TCNL
⍝            LOGICAL STATEMENT FOR VARIANCE OUTPUT
 →(D=6,7,8)/VAR
 'THE P-VALUE FOR H0: MX = MY̲ V̲E̲R̲S̲U̲S̲ H1: MX̲ ',(⍕LOGIC[D-1;1]),' MY̲ IS: ',
 (6 4 ⍕PVM[D-1;1]),⎕TCNL
 →L8
VAR:PVM←(3,1)⍴(PV,PVI,PV3)
 'THE P-VALUE FOR H0: VX = VY̲ V̲E̲R̲S̲U̲S̲ H1: VX̲ ',(⍕LOGIC[D-5;1]),' VY̲ IS: ',
 (6 4 ⍕PVM[D-5;1]),⎕TCNL
 'PRESS ENTER WHEN READY.'
 BB←⍞
 →N3
L8:'WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE SHIFT IN LOCATION (MY̲ -
 MX̲)? (Y/N).',⎕TCNL
 BB←⍞
 →(BB='N')/N3
L10:CC←INPUT 3
 ALPHA←(100-CC)÷200
⍝      ROUTE TO NORMAL APPROX. FOR CONF. INT. OF LARGER SAMPLE SIZES
 →((NM×2)>80)/L5
⍝            COMPUTING CONFIDENCE INTERVALS BY EXACT P-VALUE
 CDF←P
⍝        INDEX POSITION OF VALUE IN CDF ≤ ALPHA
 INDEX←(+/(CDF≤ALPHA))
 →(INDEX>0)/L6
 INDEX←1
 →L6
⍝         COMPUTING CONFIDENCE INTERVALS USING NORMAL APPROX. W/C.F.
L5:DEN←((MM×NN×(MM+NN+1))÷12)*0.5
 UALPHA←(DEN×(NORMPTH ALPHA))+NM-0.5
⍝             ROUND UALPHA DOWN AND INCREMENT BY ONE
```

```
  INDEX←⌊UALPHA+1
 L6:IPX←NN,INDEX
  CI←IPX CONFMW A
  '      A ',(⍕CC),'  CONFIDENCE INTERVAL FOR THE SHIFT IN '
  'LOCATION BETWEEN POPULATIONS X AND Y IS:',⎕TCNL
  '                  ( ',(⍕CI[1]),' ≤ MY - MX ≤ ',(⍕CI[2]),' )',⎕TCNL
  'PRESS ENTER WHEN READY.'
  BB←⎕
  →N3
 E1:'ERROR: SAMPLE CONTAINS LESS THAN TWO ENTRIES; TRY AGAIN.',⎕TCNL
  →B1
 E3:'ERROR: YOU HAVE ENTERED MORE THAN ONE VALUE. TRY AGAIN.',⎕TCNL
  →B3
 ∇



 ∇ NPLR;N;SUMX;SUMY;XBAR;YBAR;SUMX2;SUMXY;B;A;WW;XX;BB;U;D;ALPHA;P;CC;CDF;
   TALPHA;NN;CI;SLOPES;RR;SR;YY;DENOM;INDEX;FF;X;Y;Q;R;CHA;E;PV
  ⍝     PROGRAM CONDUCTS NONPARAMETRIC LINEAR REGRESSION. THE LEAST SQUARES
  ⍝ ESTIMATED REGRESSION LINE IS COMPUTED WITH HYPOTHESIS TESTING AND
  ⍝ CONFIDENCE INTERVAL AVAILABLE FOR THE SLOPE B. IF B DOES NOT LIE IN THE
  ⍝ C.I. AN ALTERNATE REGRESSION LINE IS PROPOSED. SUBPROGRAMS CALLED ARE
  ⍝ SPMANP, KENDALP, NORMPTH, INPUT, AND CONFLR.
  ⎕PP←5
  ⍝ DISPLAY MENU AND INPUT DATA.
 N1:E←MENU NPLROBJ
  →(E=1)/N2
  MENU MAINOBJ
  →0
 N2:R←INPUT 2
  Q←1↑R
  X←1↓(Q+1)↑R
  Y←(Q+1)↓R
  ⍝              ASSIGN THE SIZE OF X (AND Y) TO N
  N←⍴X
  ⍝                COMPUTE THE SUM OF X'S AND Y'S
  SUMX←+/X
  SUMY←+/Y
  ⍝                COMPUTE THE MEAN OF X AND Y
  XBAR←SUMX÷N
  YBAR←SUMY÷N
  ⍝             COMPUTE THE SUM OF THE X'S SQUARED
  SUMX2←+/(X*2)
  ⍝                COMPUTE THE SUM OF X TIMES Y
  SUMXY←+/(X×Y)
  ⍝  COMPUTE 'B', THE SLOPE OF THE ESTIMATED LEAST SQUARES REGRESSION LINE
  B←((N×SUMXY)-(SUMX×SUMY))÷((N×SUMX2)-(SUMX*2))
  ⍝               COMPUTE 'A', THE Y-INTERCEPT
  A←YBAR-(B×XBAR)
  FF←'N'
  'THE LEAST SQUARES ESTIMATED REGRESSION EQUATION IS:',⎕TCNL
  'Y = ',(5⍕A),' + ',(5⍕B),'X.',⎕TCNL
  'DO YOU WISH TO ENTER SOME X VALUES TO GET THE PREDICTED Y''S? (Y/N).'
  WW←⍞
  →(WW='N')/L1
 L2:'ENTER X VALUES.'
  XX←⎕
  ⍝                 CALCULATE PREDICTED Y'S
  YY←A+B×XX
  'THE PREDICTED Y VALUES ARE: ',(⍕YY),⎕TCNL
  ' WOULD YOU LIKE TO RUN SOME MORE X VALUES? (Y/N).'
  WW←⍞
  →(WW='Y')/L2
 L1:'WOULD YOU LIKE TO TEST HYPOTHESIS ON B, THE SLOPE OF THE EQUATION? (Y/
  N).'
  WW←⍞
  →(WW='N')/L3
  'ENTER THE HYPOTHESIZED SLOPE.'
  BB←⎕
  ⍝             COMPUTE UI'S
  U←Y-(BB×X)
  CHA←'> '
  ⍝  CALL SPMANP TO COMPUTE R AND ASSOCIATED P-VALUES.
  D←X SPMANP U
  →(D[1]>0)/L11
  CHA←'< '
 L11:PV←2×D[2]
  →(PV≤1)/L19
  PV←1
 L19:'SPEARMAN''S R EQUALS: ',(4⍕D[1]),⎕TCNL
  'THE P-VALUE FOR H0: B = ',(⍕BB),' VERSUS'
  '                H1: B ',(⍕CHA[1 2]),(⍕BB),' IS: ',(4⍕D[2]),⎕TCNL
  'THE P-VALUE FOR THE TWO-SIDED TEST OF HYPOTHESIS IS: ',(4⍕(2×D[2])),⎕TCN
  L
  ⍝    IF USING THE NEW REGRESSION EQUATION BASED ON MEDIANS, EXIT HERE.
 L3:→(FF='N')/L18
  'PRESS ENTER WHEN READY.'
  WW←⍞
```

```
[69]   →N1
[70]  ⍝           COMPUTE CONFIDENCE INTERVALS ON B
[71]  L18:'WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE SLOPE? (Y/N).'
[72]   WW←⍞
[73]   →(WW='N')/N1
[74]  L10:CC←INPUT 5
[75]  ⍝      CHANGE ENTERED VALUE TO ALPHA
[76]   ALPHA←(100-CC)÷200
[77]  ⍝    ROUTE TO NORMAL APPROX. FOR CONF. INT. OF LARGER SAMPLE SIZES
[78]   →(N>13)/L5
[79]  ⍝          COMPUTING CONFIDENCE INTERVALS BY EXACT P-VALUE
[80]   P←KENDALP N
[81]   CDF←P[2;]
[82]  ⍝      INDEX POSITION OF VALUE IN CDF ≤ ALPHA
[83]   INDEX←(+/(CDF≤ALPHA))
[84]   →(INDEX>0)/L6
[85]   INDEX←1
[86]   →L6
[87]  ⍝        COMPUTING CONFIDENCE INTERVALS USING NORMAL APPROX. W/C.F.
[88]  L5:DENOM←((N×(N-1)×((2×N)+5))÷18)*0.5
[89]   TALPHA←DENOM×(|(NORMPTH ALPHA))
[90]   →L9
[91]  L6:TALPHA←P[3;INDEX]
[92]   TALPHA
[93]  L9:CI←X CONFLR Y
[94]   NN←1↑CI
[95]   SLOPES←1↓CI
[96]   RR←⌊((NN-TALPHA)÷2)
[97]   RR
[98]   →(RR=0)/L20
[99]   RR←1
[100]  L20:SR←⌈(1+((NN+TALPHA)÷2))
[101]   SR
[102]   →(SR≤(⍴SLOPES))/L21
[103]   SR←⍴SLOPES
[104]  L21:'A ',(⍕CC),'  CONFIDENCE INTERVAL FOR B, THE SLOPE OF '
[105]   ' THE ESTIMATED REGRESSION LINE, IS:',⎕TCNL
[106]   '           ( ',(5⍕SLOPES[RR]),' ≤ B ≤ ',(5⍕SLOPES[SR]),' ).',⎕TCNL
[107]   'PRESS ENTER WHEN READY.'
[108]   WW←⍞
[109]  ⍝ IF B OUTSIDE THE C.I. CALCULATE NEW EQUATION BASED ON MEDIANS
[110]   →((B≥SLOPES[RR])∧(B≤SLOPES[SR]))/N1
[111]  ⍝                      ORDER X AND Y
[112]   X←X[⍋X]
[113]   Y←Y[⍋Y]
[114]  ⍝     CHECK TO SEE IF THE SIZE OF SS IS EVEN OR ODD FOR FINDING MEDIANS
[115]   →((2|NN)=0)/S1
[116]  ⍝                COMPUTE MEDIAN FOR ODD CASE
[117]   B←SLOPES[((NN+1)÷2)]
[118]   →S2
[119]  ⍝                COMPUTE MEDIAN FOR EVEN CASE
[120]  S1:B←(SLOPES[(NN÷2)]+SLOPES[((NN+2)÷2)])÷2
[121]  ⍝       DO THE SAME FOR THE X AND Y VECTORS
[122]  S2:→((2|N)=0)/S3
[123]   YBAR←Y[((N+1)÷2)]
[124]   XBAR←X[((N+1)÷2)]
[125]   →OUT
[126]  S3:YBAR←(Y[(N÷2)]+Y[((N+2)÷2)])÷2
[127]   XBAR←(X[(N÷2)]+X[((N+2)÷2)])÷2
[128]  ⍝                 COMPUTE NEW INTERCEPT 'A'
[129]  OUT:A←YBAR-(B×XBAR)
[130]   'THE LEAST SQUARES ESTIMATOR OF B LIES OUTSIDE THE CONFIDENCE INTERVAL.'
[131]   'DISCARD THE LEAST SQUARES EQUATION AND USE:',⎕TCNL
[132]   '                  Y = ',(⍕A),' + ',(⍕B),'X ',⎕TCNL
[133]   '         THIS EQUATION IS BASED ON THE MEDIANS OF THE X AND Y DATA,AND TH
        E'
[134]   'MEDIAN OF THE TWO-POINT SLOPES CALCULATED FOR THE CONFIDENCE INTERVAL O
        N B.',⎕TCNL
[135]  ⍝    ALLOW USER TO DO SOME ANALYSIS ON NEW EQUATION
[136]   '         DO YOU WISH TO ENTER SOME X VALUES TO GET PREDICTED Y''S '
[137]   'FROM THE NEW EQUATION? (Y/N).'
[138]   FF←⍞
[139]   →(FF='Y')/L2
[140]   FF←'Y'
[141]   →L1
      ∇


      ∇ SIGN;A;C;B;D;PVAL;X;MO;N;CDF;ALPHA;CI;Y;AA;BB;CC;DD;PV;PVI;NNN;KPOS;ORD
        D;YY;ORDX;KALPHA;Z;Z1;QUA;WW;PVM;PV3;R;Q
[1]   ⍝      THIS FUNCTION USES THE ORDINARY SIGN TEST TO CALCULATE THE K
[2]   ⍝ STATISTIC, P-VALUE, AND CONFIDENCE INTERVAL AS A TEST FOR MEDIANS.
[3]   ⍝ THE LAST OPTION WILL DISPLAY A TABLE OF CONFIDENCE INTERVALS OF ORDERED
[4]   ⍝ STATISTICS WITH CONFIDENCE COEFFICIENTS.
[5]   ⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: BINOM, NORMCDF, NORMPTH,
[6]   ⍝ INPUT,AND QUANC.
[7]    →N4
[8]   N1:MENU SIGNHELP
[9]   ⍝       MENU CHOICES AND ROUTE FOR PROPER ACTIONS
[10]  N4:C←CHOICES PAGEDMENU SIGNOBJ
```

```
 →(C=2,3,4,5)/L8
 →(C=7,8,9,10)/L9
 →(C=1)/N1
 →(C=11)/L20
 MENU MAINOBJ
 →0
⍝          INPUT DATA FOR SINGLE SAMPLE CASE
L8:AA←1
 X←INPUT 1
 NNN←⍴X
 →(C=5)/L16
 MO←INPUT 3
 D←X-MO
 →L11
⍝                         PAIRED SAMPLE CASE
L9:AA←2
 R←INPUT 2
 Q←1↑R
 X←1↓(Q+1)↑R
 Y←(Q+1)↓R
 DD←X-Y
 NNN←⍴DD
 →(C=10)/L16
 MO←INPUT 4
 D←(X-Y)-MO
⍝                         COMPRESS D TO REMOVE ZEROS
L11:A←(D≠0)/D
⍝                   RECORD LENGTH OF A AND ASSIGN TO N
 N←⍴A
⍝                        KEEPING TRACK OF POSITIVE SIGNS
 KPOS←+/(A>0)
 →(N>25)/NORM
 PVAL←BINOM N
 →(KPOS>0)/P1
 PVI←1
 →P2
P1:PVI←1-PVAL[KPOS]
P2:PV←PVAL[(KPOS+1)]
 →L6
⍝ IF N IS GREATER THAN 30 USE NORMAL APPROX W/ CONTINUITY CORRECTION
NORM:Z1←((KPOS-0.5)-(0.5×N))÷(0.5×(N*0.5))
 Z←((KPOS+0.5)-(0.5×N))÷(0.5×(N*0.5))
 PV←NORMCDF Z
 PVI←1-(NORMCDF Z1)
⍝            IF PAIRED SAMPLE TEST GO TO L17 FOR OUTPUT STATEMENT
L6:PV3←2×(⌊/(PV,PVI))
 →(PV3≤1)/N5
 PV3←1
N5:PVM←(3,1)⍴(PV,PVI,PV3)
 'COMPUTATIONS ARE BASED ON A SAMPLE SIZE OF: ',(⍕N),'.',⎕TCNL
 'THE TOTAL NUMBER OF POSITIVE SIGNS IS: ',(⍕KPOS),⎕TCNL
 →(C=7,8,9)/L17
 'THE P-VALUE FOR H0: M = ',(⍕MO),' VERSUS H1: M ',(⍕LOGIC[(C-1);1]),' ',(
 ⍕MO),' IS: ',(5⍕PVM[C-1;1]),⎕TCNL
 →L18
L17:'THE P-VALUE FOR COMPARING THE MEDIAN OF THE POPULATION OF '
 'DIFFERENCES TO THE HYPOTHESIZED MEDIAN, ',⎕TCNL
 'H0: M(X-Y) = ',(⍕MO),' VERSUS H1: M(X-Y) ',(⍕LOGIC[(C-6);1]),' ',(⍕MO),'
 , IS: ',(5⍕PVM[(C-6);1]),⎕TCNL
L18:'WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE MEDIAN? (Y/N).',⎕TCNL
 BB←⍞
 →(BB='Y')/L16
 →QUANT
⍝                 INPUT SIZE OF CONFIDENCE INTERVAL
L16:CC←INPUT 5
 ALPHA←(100-CC)÷200
 →(NNN>25)/NORM1
⍝        COMPUTING CONFIDENCE INTERVALS BY EXACT P-VALUE
 CDF←BINOM NNN
⍝               INDEX POSITION OF CDF FOR ALPHA ÷ 2
 B←+/(CDF≤ALPHA)
 →(B>0)/SKIP
 B←1
 →SKIP
⍝ COMPUTING CONFIDENCE INTERVALS BY NORMAL APPROX.
NORM1:KALPHA←((0.5×(NNN*0.5))×(NORMPTH ALPHA))+(0.5×NNN)-0.5
⍝  ROUND KALPHA DOWN TO NEAREST INTEGER AND INCREMENT BY ONE
 B←⌊KALPHA+1
⍝                    IF SINGLE SAMPLE CASE GO TO L7
SKIP:→(C=2,3,4,5)/L7
⍝               CALCULATE AND PRINT OUT CONF. INT. FOR PAIRED SAMPLE CASE
L5:ORDD←DD[⍋DD]
 YY←⍴ORDD
 CI←ORDD[B],ORDD[(YY-(B-1))]
 '        A ',(⍕CC),'   CONFIDENCE INTERVAL FOR THE MEDIAN OF THE '
 'POPULATION OF DIFFERENCES IS:',⎕TCNL
 '                      ( ',(⍕CI[1]),' ≤ MEDIAN(X-Y) ≤ ',(⍕CI[2]),' )',⎕TCNL
 →QUANT
⍝               CALCULATE AND PRINT OUT CONF. INT. FOR ONE SAMPLE CASE
L7:ORDX←X[⍋X]
 YY←⍴ORDX
```

```
[100]  CI←ORDX[B],ORDX[(YY-(B-1))]
[101]  'A ',(⍕CC),'  CONFIDENCE INTERVAL FOR THE MEDIAN OF THE POPULATION IS: '
       ,⎕TCNL
[102]  '                    ( ',(⍕CI[1]),' ≤ MEDIAN ≤ ',(⍕CI[2]),' )',⎕TCNL
[103] QUANT:'WOULD YOU LIKE CONFIDENCE INTERVALS FOR A SPECIFIED QUANTILE? (Y/N
       ).',⎕TCNL
[104]  WW←⍞
[105]  →(WW='Y')/B1
[106]  →N4
[107] L20:'ENTER THE SIZE OF THE SAMPLE.'
[108]  NNN←⎕
[109] B1:'ENTER DESIRED QUANTILE; FOR EXAMPLE: ENTER 20, FOR THE 20TH QUANTILE.
       '
[110]  QUA←⎕
[111]  →((QUA≤0)∨(QUA>100))/E1
[112]  NNN QUANC QUA
[113]  ' '
[114]  ' ***** THIS TABLE GIVES CONFIDENCE COEFFICIENTS FOR VARIOUS INTERVALS '
[115]  'WITH ORDER STATISTICS AS END POINTS FOR THE ',(⍕QUA),'TH QUANTILE.',⎕TC
       NL
[116]  'PRESS ENTER WHEN READY.'
[117]  BB←⍞
[118]  →N4
[119] E1:'ERROR: THE QUANTILE VALUE MUST LIE BETWEEN 0 AND 100; TRY AGAIN.'
[120]  →B1
      ∇


      ∇ SPMAN;X;Y;A;Q;R;CHA;BB;B;PV
[1]   ⍝      THIS FUNCTION COMPUTES THE SPEARMAN R STATISTIC WHICH MEASURES
[2]   ⍝ THE DEGREE OF CORRESPONDENCE BETWEEN RANKINGS OF TWO SAMPLES. THE P-
[3]   ⍝VALUE IS GIVEN FOR TESTING ONE AND TWO-SIDED HYPOTHESIS OF ASSOCIATION
[4]   ⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: TIES, TIESK, SPEARP,
[5]   ⍝ INPUT, SPAPROX, INTERP, AND THE VARIABLE PMATSP.
[6]
[7]   ⍝   DISPLAY MENU AND INPUT DATA.
[8]   N1:B←MENU SPMANOBJ
[9]    →(B=1)/B1
[10]   MENU MAINOBJ
[11]   →0
[12]  B1:R←INPUT 2
[13]   Q←1↑R
[14]   X←1↓(Q+1)↑R
[15]   Y←(Q+1)↓R
[16]  ⍝ CALL SPMANP TO CALCULATE THE STATISTIC AND ASSOCIATED P-VALUES
[17]   A←X SPMANP Y
[18]   →(A[1]>0)/L1
[19]   CHA←'INDIRECT'
[20]   →L2
[21]  L1:CHA←'DIRECT'
[22]  L2:PV←2×A[2]
[23]   →(PV≤1)/L3
[24]   PV←1
[25]  L3:'SPEARMAN''S R EQUALS: ',(4⍕A[1]),⎕TCNL
[26]   'THE P-VALUE FOR H0: NO ASSOCIATION EXISTS VERSUS'
[27]   '                 H1: ',(⍕CHA),' ASSOCIATION EXISTS IS: ',(4⍕A[2]),⎕TCNL
[28]   'THE P-VALUE FOR THE TWO-SIDED TEST OF HYPOTHESIS IS: ',(4⍕(2×A[2])),⎕TCN
       L
[29]   'PRESS ENTER WHEN READY.'
[30]   BB←⍞
[31]   →N1
      ∇


      ∇ WISIG;A;B;D;E;F;PV2;Z1;Z;DEN;NUMZ;NUMZ1;PVAL;X;MO;N;TPLUS;CDF;TALPHA;AL
        PHA;H;CI;Y;AA;BB;CC;NN;DD;PV;POS;TPOS;NM;PVI;TPOS1;NNN;C;PVM;PV3;R;Q;NU
        M;TC;TC1;TRAP;DENT;DENT1
[1]   ⍝ THIS FUNCTION USES THE WILCOXON SIGNED RANK TEST TO CALCULATE THE TPLUS
[2]   ⍝ STATISTIC, P-VALUE, AND CONFIDENCE INTERVAL AS A TEST FOR MEDIANS.
[3]   ⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: TIES, WILP, NORMCDF,
[4]   ⍝ NORMPTH, CONFW, INPUT AND THE VARIABLE PMATRIX.
[5]    →N4
[6]   N1:MENU WILHELP
[7]   ⍝          MENU CHOICES AND ROUTE FOR PROPER ACTIONS.
[8]   N4:C←CHOICEW PAGEDMENU WILOBJ
[9]    →(C=2,3,4,5)/L8
[10]   →(C=6,7,8,9)/L9
[11]   →(C=1)/N1
[12]   MENU MAINOBJ
[13]   →0
[14]  ⍝                    INPUT DATA FOR SINGLE SAMPLE CASE
[15]  L8:AA←1
[16]   X←INPUT 1
[17]   NNN←⍴X
[18]   →(C=5)/L16
[19]   MO←INPUT 3
```

```
 D←X-MO
 →L11
⍝                         PAIRED SAMPLE CASE
L9:AA←2
 R←INPUT 2
 Q←1↑R
 X←1↓(Q+1)↑R
 Y←(Q+1)↓R
 DD←X-Y
 NNN←⍴DD
 →(C=9)/L16
 MO←INPUT 4
 D←(X-Y)-MO
⍝                         COMPRESS D TO REMOVE ZEROS
L11:A←(D≠0)/D
⍝                     RECORD LENGTH OF A AND ASSIGN TO N
 N←⍴A
⍝                       KEEPING TRACK OF POSITIVE SIGNS
 POS←(A>0)
⍝             TAKE THE ABSOLUTE VALUE OF A; ASSIGN TO B AND ORDER B
 B←|A
 B←B[⍋B]
⍝      REORDER POSITIVE SIGNS TO COINCIDE WITH PROPER POSITIONS IN B
 POS←POS[⍋(|A)]
⍝                         CALL FUNCTION TO BREAK TIES
 E←1 TIES B
⍝      CALCULATE TPOS BY ADDING ACROSS ALL POSITIVE VALUES OF E
 TPOS←+/(POS×E)
 TPOS1←TPOS
⍝         GIVES SIZE OF LEFT TAIL OF PROBABILITY DISTRIBUTION
 NM←(⌊((+/⍳N)÷2))+1
⍝             GO TO STATEMENTS BASED ON LENGTH OF VECTOR E
 →(N>9)/L3
⍝                   GENERATE NULL DISTRIBUTION FOR TPLUS
 F←WILP N
⍝  IF TPOS FALLS IN LEFT HALF OF PROB DIST CALCULATE PVALUE AS NORMAL
L1:→(TPOS≤(NM-1))/TREG
⍝         OTHERWISE USE THE NEGATIVE T STATISTIC
 TPOS←(+/⍳N)-TPOS
⍝ IF TPOS IS FRACTIONAL USE BOTH THE INTEGER ABOVE AND BELOW AS TPLUS
TREG:→((1|TPOS)=0)/NON
 TPLUS←⌊TPOS
P1:PV←1-((F[TPLUS]+F[TPLUS+1])÷2)
 PVI←(F[TPLUS+1]+F[TPLUS+2])÷2
 →CHECK
NON:→(TPOS>0)/GO
 PV←1
 →P2
GO:PV←1-F[(TPOS)]
P2:PVI←F[(TPOS+1)]
CHECK:→(TPOS1≤(NM-1))/L6
 PV2←PV
 PV←PVI
 PVI←PV2
 →L6
⍝  COMPUTE NORMAL APPROX. W/CONTINUITY CORRECTION FACTOR
L3:TRAP←(N×(N+1))÷4
 NUMZ←(TPOS+0.5)-TRAP
 NUMZ1←(TPOS-0.5)-TRAP
 DEN←((N×(N+1)×((2×N)+1))÷24)*0.5
 Z←NUMZ÷DEN
 Z1←NUMZ1÷DEN
⍝ COMPUTE STUDENT T APPROXIMATION WITH CONTINUITY CORRECTION FACTOR
 NUM←|(TPOS-TRAP)
 DENT←(((N×(DEN*2))÷(N-1))-(((NUM-0.5)*2)÷(N-1)))*0.5
 DENT1←(((N×(DEN*2))÷(N-1))-(((NUM+0.5)*2)÷(N-1)))*0.5
 TC←(NUM-0.5)÷DENT
 TC1←(NUM+0.5)÷DENT1
⍝ COMPUTE AVERAGE OF TC AND ZC
 →(TPOS≤((+/⍳N)÷2))/SECOND
 PV←((1-(NORMCDF Z1))+(1-((N-1) TDISTN TC1)))÷2
 PVI←((NORMCDF Z)+((N-1) TDISTN TC))÷2
 →L6
SECOND:PV←((1-(NORMCDF Z1))+((N-1) TDISTN TC1))÷2
 PVI←((NORMCDF Z)+(1-((N-1) TDISTN TC)))÷2
L6:PV3←2×(⌊/(PV,PVI))
 →(PV3≤1)/N5
 PV3←1
N5:PVM←(3,1)⍴(PVI,PV,PV3)
 'COMPUTATIONS ARE BASED ON A SAMPLE SIZE OF ',(⍕N),'.',⎕TCNL
 'THE TOTAL SUM OF POSITIVE RANKS IS: ',(⍕TPOS1),'.',⎕TCNL
⍝          IF PAIRED SAMPLE TEST GO TO L17 FOR OUTPUT STATEMENT
 →(AA=2)/L17
 'THE P-VALUE FOR H0: M = ',(⍕MO),' VERSUS H1: M ',(⍕LOGIC[C-1;1]),' ',(⍕
 MO),' IS:',(4⍕PVM[C-1;1]),⎕TCNL
 →L18
L17:'     THE P-VALUE FOR COMPARING THE MEDIAN OF THE POPULATION OF '
 'DIFFERENCES TO THE HYPOTHESIZED MEDIAN, ',⎕TCNL
 'H0: M(X-Y) = ',(⍕MO),' VERSUS H1: M(X-Y) ',(⍕LOGIC[C-5;1]),' ',(⍕MO),',
  IS:',(4⍕PVM[C-5:1]).⎕TCNL
```

```
[108] L18:'WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE MEDIAN? (Y/N).'
[109]  BB←⎕
[110]  →(BB='Y')/L16
[111]  →N4
[112] L16:CC←INPUT 5
[113]  ALPHA←(100-CC)÷200
[114] ⍝          ROUTE TO NORMAL APPROX. FOR CONF INT OF LARGE SAMPLE SIZE
[115]  →(NNN>9)/L4
[116]  CDF←WILP NNN
[117] ⍝                INDEX POSITION OF CDF FOR ALPHA ÷ 2
[118]  CDF←(CDF≠0)/CDF
[119]  TALPHA←(+/(CDF≤ALPHA))
[120]  →(TALPHA≠0)/JUMP
[121]  TALPHA←1
[122]  →JUMP
[123] ⍝ COMPUTING CONFIDENCE INTERVALS BY NORMAL APPROX. W/C.F.
[124] L4:TRAP←(NNN×(NNN+1))÷4
[125]  DEN←((NNN×(NNN+1)×((2×NNN)+1))÷24)*0.5
[126]  TALPHA←(DEN×(NORMPTH ALPHA))+TRAP-0.5
[127]  TALPHA
[128] ⍝          ROUND TALPHA DOWN TO INTEGER VALUE AND INCREMENT BY ONE
[129]  TALPHA←⌊TALPHA+1
[130] ⍝                IF ONE SAMPLE CASE GO TO L7
[131] JUMP:→(C=2,3,4,5)/L7
[132] ⍝              CALCULATE AND PRINT OUT CONF. INT. FOR PAIRED SAMPLE CASE
[133] L5:CI←TALPHA CONFW DD
[134]  '     A ',(⍕CC),'  CONFIDENCE INTERVAL FOR THE MEDIAN OF THE '
[135]  'POPULATION OF DIFFERENCES IS:',⎕TCNL
[136]  '              ( ',(⍕CI[1]),' ≤ MEDIAN(X-Y) ≤ ',(⍕CI[2]),' )',⎕TCNL
[137]  'PRESS ENTER WHEN READY.'
[138]  BB←⎕
[139]  →N4
[140] ⍝              CALCULATE AND PRINT OUT CONF. INT. FOR ONE SAMPLE CASE
[141] L7:CI←TALPHA CONFW X
[142]  'A ',(⍕CC),'  CONFIDENCE INTERVAL FOR THE MEDIAN OF THE POPULATION IS:',,
       ⎕TCNL
[143]  '              ( ',(⍕CI[1]),' ≤ MEDIAN ≤ ',(⍕CI[2]),' )',⎕TCNL
[144]  'PRESS ENTER WHEN READY.'
[145]  BB←⎕
[146]  →N4
      ∇
```

APPENDIX E

MAIN PROGRAM LISTINGS FOR MAINFRAME COMPUTER WORKSPACE

```
  ∇ KENDALL;A;AA;B;BX;BY;C;CX;CY;D;DD;DX;DY;DXY;S;POS;NEG;XX;YY;N;DEN;NN;NU
    M;P;PV;PVAL;SU;SV;T;U;V;AT;Z;X;Y;CHA;Q;R
⍝   THIS FUNCTION COMPUTES THE KENDALL B STATISTIC WHICH IS A MEASURE
⍝ OF ASSOCIATION BETWEEN SAMPLES. P-VALUES ARE GIVEN FOR TESTING ONE
⍝ AND TWO-SIDED HYPOTHESIS FOR NO ASSOCIATION VERSUS ASSOCIATION.
⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: TIES, TIESK, KENDALP,
⍝ INTERP INPUT AND NORMCDF.
 R←INPUT 2
 Q←1↑R
 X←1↓(Q+1)↑R
 Y←(Q+1)↓R
⍝             ORDER Y IN INCREASING ORDER OF X
 A←Y[⍋X]
⍝            ORDER X IN INCREASING ORDER
 B←X[⍋X]
⍝              COMPUTE CURRENT RANKING OF Y
 C←⍋⍋A
⍝              NOW ORDER Y RANKS IN INCREASING ORDER
 D←A[⍋A]
⍝     IF TIES EXIST IN EITHER X OR Y RANKED VECTOR USE MID-RANK METHOD
 DD←1 TIES D
 XX←1 TIES B
⍝        FIND ORIGINAL RANKING OF Y WITH TIES RESOLVED
 YY←DD[C]
 N←⍴X
⍝ COMPUTE NUMBER OF DISTINGUISHABLE PAIRS
 NN←(N×(N-1))÷2
 S←⍴0
 AA←0
⍝ POSITIVE ONES COME FROM A RUNS UP CONDITION; NEGATIVE 1 FROM RUNS DOWN
⍝ ZERO IS SCORED FOR TIES. MULTIPLY THE RESULTS FOR EVERY ELEMENT AND SUM
L1:AA←AA+1
 BX←(XX[AA]>(AA↓XX))
 CX←(XX[AA]<(AA↓XX))×(¯1)
 DX←BX+CX
 BY←(YY[AA]>(AA↓YY))
 CY←(YY[AA]<(AA↓YY))×(¯1)
 DY←BY+CY
 DXY←DX×DY
 POS←(DXY>0)
 NEG←(DXY<0)×(¯1)
 S←S,POS,NEG
 →(AA<(N-1))/L1
⍝            SUM FINAL VECTOR TO DETERMINE S
 S←+/S
⍝    OBTAIN THE NUMBER OF TIES IN EACH VECTOR USING THE TIESK FUNCTION
 U←TIESK B
 V←TIESK D
 SU←+/(2!U)
 SV←+/(2!V)
⍝ CALCULATE THE B STATISTIC INCLUDING THE CORRECTION FOR TIES
 T←S÷((NN-SU)×(NN-SV))*0.5
 AT←|T
 →(N>12)/NORM
⍝       CALL KENDALP TO CALCULATE THE RIGHT TAIL OF THE CDF OF B
 P←KENDALP N
⍝ CALL INTERP TO CALCULATE P-VALUE BY INTERPOLATION
 PVAL←AT INTERP P
 →(PVAL≠¯1)/L3
 PVAL←0.5
 →L3
⍝ CALCULATE P VALUE USING NORMAL APPROX.
NORM:NUM←(3×AT)×((2×NN)*0.5)
 DEN←(2×((2×N)+5))*0.5
 Z←NUM÷DEN
 PVAL←1-(NORMCDF Z)
⍝ IF B IS POSITIVE PRINT OUT DIRECT ASSOCIATION.
L3:→(T>0)/L5
 CHA←'INDIRECT'
 →L7
L5:CHA←'DIRECT'
L7:PV←2×PVAL
 →(PV≤1)/L8
 PV←1
L8:'KENDALL''S B EQUALS ',(4⍕T)
 ' '
 'THE P-VALUE FOR HO: NO ASSOCIATION EXISTS VERSUS'
 '                 H1: ',(⍕CHA),'ASSOCIATION EXISTS IS: ',(4⍕PVAL)
 ' '
 'THE P-VALUE FOR THE TWO-SIDED TEST OF HYPOTHESIS IS:',(4⍕PV)
 ' '
 ∇
```

```
 ∇ KRUSKAL;NUM;DENOM;A;C;H;D;K;AA;BB;DD;E;F;N;OF;P;PVAL;R;SOFR;SR;TSOR;CHA
⍝    THIS FUNCTION COMPUTES THE KUSKAL-WALLIS TEST STATISTIC H WHICH IS
⍝ A MEASURE OF THE EQUALITY OF K INDEPENDENT SAMPLES.
⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: TIES, TIESK, INDEXPLS,
⍝ FDISTN, INTERP AND THE VARIABLES PMATKW20, PMATKW31, PMATKW33,
⍝ PMATKW34, PMATKW41, PMATKW42, AND PMATKW43.
 ⎕PP←5
E1:'ENTER THE NUMBER OF POPULATIONS TO BE COMPARED (MUST BE GREATER THAN T
 WO).'
 K←⎕
 →((K<3)∨((1|K)=0))/E1
 →((⍴K)>1)/E1
⍝                INITIALIZE VECTORS E AND F AND VARIABLE C
 E←F←SOFR←⍳0
 C←0
⍝THIS LOOP FACILITATES ENTERING THE SAMPLE VECTORS AND STORING THEM
 CHA←'FIRST'
L1:C←C+1
 'ENTER YOUR ',(⍕CHA),' SAMPLE.'
 D←⎕
 →((⍴⍴D)=0)/NEXT
 D←1⍴D
⍝   CONCATENATE SAMPLES AS THEY ARE ENTERED AND STORE THEM IN VECTOR E
NEXT:E←E,D
⍝          RECORD THE LENGTHS OF THE SAMPLES AS THEY ARE ENTERED
 F←F,⍴D
 CHA←'NEXT'
 →(C<(K-1))/L1
 CHA←'LAST'
 →(C<K)/L1
⍝        RECORD SIZE OF ALL SAMPLES WHEN COMBINED
 N←+/F
⍝ ORDER SAMPLE SIZES LARGEST TO SMALLEST
 OF←F[⍒F]
⍝          ORDER COMBINED SAMPLE VECTOR TO BE USED BY TIES FUNCTION
 D←E[⍋E]
⍝   CALL INDEXPLS TO INCREMENT INDEXES WHEN TIES OCCUR WITHIN ONE SAMPLE
 AA←F INDEXPLS E
⍝           CALL TIES TO BREAK TIES BY MIDRANK METHOD
 BB←1 TIES D
 C←0
⍝             THIS LOOP CALCULATES THE H STATISTIC
L2:C←C+1
⍝             SUM OF RANKS FOR EACH SAMPLE IS CALCULATED
 SR←+/BB[(F[C]↑(AA[C;]))]
⍝  CALCULATE SUM OF RANKS SQUARED DIVIDED BY THE INDIVIDUAL SAMPLE SIZE
 SR←(SR*2)÷F[C]
⍝              STORE EACH CALCULATION
 SOFR←SOFR,SR
 →(C<K)/L2
⍝              SUM ACROSS ALL SAMPLES
 TSOR←+/SOFR

 H←(TSOR×(12÷(N×(N+1))))-(3×(N+1))
⍝  RECALCULATE H WITH CORRECTION FOR TIES
 A←TIESK E
 NUM←(+/(A*3))-(+/A)
 DENOM←N×((N*2)-1)
 H←H÷(1-(NUM÷DENOM))
⍝  SYSTEM OF LOGICAL STATEMENTS ENSURE PROPER PROB. IS ACCESSED
 →(OF[1]<2)/OUTPUT
 →(K>4)/FAPPROX
 →(K=3)/IF
 →((∧/(OF= 3 3 3 3))∨(OF[1]>3))/FAPPROX
 →(∧/(OF= 2 1 1 1))/OUTPUT
 →((K=4)∧(OF[1]=2))/P42
 →((∧/(OF= 3 1 1 1))∨(∧/(OF= 3 2 1 1))∨(∧/(OF= 3 3 1 1))∨(∧/(OF= 3 3 2 1))
 )/P41
 →(K=4)/P43
IF:→(OF[1]>4)/FAPPROX
 →((∧/(OF= 2 1 1))∨(∧/(OF= 3 1 1)))/OUTPUT
 →((OF[1]=4)∧(OF[3]=1))/P31
 →((∧/(OF= 3 2 1))∨(∧/(OF= 3 3 1))∨(∧/(OF= 3 3 2))∨(∧/(OF= 3 3 3)))/P33
 →((∧/(OF= 3 2 2))∨(OF[1]=4))/P34
⍝                  CALL APPROPRIATE VARIABLE ACCESS CDF
P23:P←PMATKW20[(N-4);;]
 →PM
P31:P←PMATKW31[(N-5);;]
 →PM
P33:P←PMATKW33[(N-5);;]
 →PM
```

```
P34:P←PMATKW34[(N-6);;]
 →PM
P41:P←PMATKW41[(N-5);;]
 →PM
P42:P←PMATKW42[(N-5);;]
 →PM
P43:P←PMATKW43[(N-7);;]
⍝     CALL INTERP TO CALCULATE P-VALUE BY INTERPOLATION
PM:PVAL←H INTERP P
 →(PVAL=¯1)/OUTPUT
 →L5
⍝  CALCULATE P-VALUE USING THE F DIST W/ONE LESS D.F.IN DENOM APPROX.
FAPPROX:F←((N-K)×H)÷((K-1)×((N-1)-H))
 PVAL←1-(((K-1),((N-K)-1)) FDISTN F)
 →L5
OUTPUT:PVAL←'GREATER THAN .25'
L5:'THE H STATISTIC EQUALS: ',(4⍕H)
 ' '
 'THE P-VALUE FOR H0: THE POPULATION MEDIANS ARE EQUAL VERSUS '
 ' '
 '      H1: AT LEAST TWO POPULATION MEDIANS ARE NOT EQUAL IS: ',(⍕PVAL)
 ' '
 →0
E1:'ERROR: YOU MUST ENTER A SINGLE INTEGER VALUE GREATER THAN 2; TRY AGAI
 N.'
 ' '
 →B1
∇


∇ MANNWHIT;N;M;PV2;A;B;C;G;MM;NN;RX;U;NM1;P;NU;PVAL;NM;NUMZ;NUMZ1;DENC1;Z
  ;Z1;ALPHA;CDF;INDEX;IPX;CI;UALPHA;BB;CC;U1;U2;PV;NN1;NN2;PVI;DIFF;AA;GG
  ;PVM;PV3;D;Q;R;DEN;DENC;TC;TC1;NUM
⍝    THIS FUNCTION USES THE SUM OF RANKS PROCEDURE TO CALCULATE THE
⍝ MANN-WHITNEY U STATISTIC USED IN COMPUTING THE P-VALUE FOR THE TEST
⍝ OF LOCATION AND SCALE.  THE C.I. FOR M(Y)-M(X), THE SHIFT IN
⍝ LOCATION, IS ALSO COMPUTED. SUBPROGRAMS CALLED INCLUDE: TIES, TIES2
⍝ INDEXPLS, VARMW, MANWP, INPUT, CONFMW, NORMCDF, AND NORMPTH.
 DIFF←0
 ' '
 'DO YOU WISH TO COMPARE THE MEDIANS OR VARIANCES OF THE POPULATIONS?'
 ' '
B2:'   ENTER:     1 TO COMPARE MEDIANS;     2 TO COMPARE VARIANCES.'
 AA←⎕
 →((AA≠1)∧(AA≠2))/E2
 →(AA=1)/N1
 '   THE TEST TO COMPARE VARIANCES REQUIRES THE TWO POPULATION MEANS'
 'OR MEDIANS TO BE EQUAL.  IF  THEY DIFFER  BY A KNOWN AMOUNT, '
 'THE DATA CAN BE ADJUSTED BEFORE APPLYING THE TEST.'
 ' '
B3:'     ENTER THE DIFFERENCE OF MEDIANS (M(X) - M(Y)) OR 900 TO QUIT).'
 DIFF←⎕
 →((⍴DIFF)>1)/E3
 →(DIFF=900)/0
 'THE NULL HYPOTHESIS STATES - THE POPULATION VARIANCES ARE EQUAL; V(X) =
 V(Y).'
 ' '
 '                    WHICH ALTERNATIVE DO YOU WISH TO TEST?'
 ' '
N2:'ENTER:  1 FOR H1: V(X) < V(Y);  2 FOR H1: V(X) > V(Y);  3 FOR H1: V(X)
  ≠ V(Y).'
 D←⎕
 →((D≠1)∧(D≠2)∧(D≠3))/E4
 →B1
N1:'  THE NULL HYPOTHESIS STATES - THE MEDIANS OF X AND Y ARE EQUAL; M(X)
 = M(Y).'
 ' '
 '                    WHICH ALTERNATIVE DO YOU WISH TO TEST?'
 ' '
B6:'ENTER:  1 FOR H1: M(X) < M(Y);  2 FOR H1: M(X) > M(Y);  3 FOR H1: M(X)
  ≠ M(Y).'
 D←⎕
 →((D≠1)∧(D≠2)∧(D≠3))/E4
⍝                     ENTER DATA VECTORS
B1:'ENTER X DATA (MORE THAN ONE OBSERVATION IS REQUIRED).'
 N←⎕
 →((⍴⍴N)=0)/E1
 'ENTER Y DATA.'
 M←⎕
⍝ IF CALCULATIONS INVOLVE VARIANCES ADJUST X BY THE DIFFERENCE IN MEANS
 N←N-DIFF
⍝         CONCATENATE X AND Y SAMPLE VECTORS
 A←N,M
⍝         DETERMINE SIZE OF X AND Y VECTORS AND ASSIGN TO NN AND MM
 NN←⍴N
 MM←⍴M
⍝         COMPUTE SIZE LIMIT OF LEFT TAIL OF NULL DISTRIBUTION
 NM←(NN×MM)÷2
 NM1←⌊NM
```

```
⍝               ORDER A AND ASSIGN TO B
 B←A[⍋A]
 C←(NN,MM) INDEXPLS A
⍝      CALL TIES FUNCTION TO BREAK TIES USING MIDRANK METHOD
 G←1 TIES B
⍝ IF FALSE CALCULATE TEST FOR VARIANCES
 →(AA=1)/B5
⍝     CALL VARMW TO GENERATE RANKS REQUIRED FOR VARIANCE TEST
 GG←VARMW(NN+MM)
⍝ CALL TIES TO RECORD TIES IN THE DATA AND BREAK TIES IN GG
 G←GG TIES B
⍝                   CALCULATE SUM OF X RANKS
B5:RX←+/(G[(C[1;(⍳NN)])])
⍝                  CONVERT TO MANNWHIT U STATISTIC
 U←RX-((NN×(NN+1))÷2)
 U1←U
⍝        IF SIZE OF X TIMES SIZE OF Y > 80; GO TO NORMAL APPROX
 →((NM×2)>80)/L2
 NN1←⌊/NN,MM
 NN2←⌈/NN,MM
⍝MANWP FUNCTION CALCULATES LEFT TAIL CUMULATIVE PROBS. OF U STATISTIC
 P←NN2 MANWP NN1
⍝     LOGICAL STATEMENT ENSURES ONLY LEFT SIDE OF NULL DIST IS USED
 →(U≤NM1)/L3
⍝      CONVERT U STAT WHEN GREATER THAN LEFT TAIL VALUES
 U←(NN×MM)-U
⍝    IF U1 IS A FRACTIONAL, INTERPOLATE P VALUE
L3:→((1|U)=0)/NON
 U2←⌊U
 →(U2>0)/P1
 PV←1-((P[U2+1])÷2)
 →P3
P1:PV←1-((P[U2]+P[U2+1])÷2)
P3:PVI←(P[U2+1]+P[U2+2])÷2
 →CHECK
NON:→(U>0)/GO
 PV←1
 →P2
GO:PV←1-P[U]
P2:PVI←P[(U+1)]
CHECK:→(U1≤NM1)/L4
 PV2←PV
 PV←PVI
 PVI←PV2
 →L4
⍝ COMPUTE THE  NORMAL APPROXIMATION W/CORRECTION FACTOR
L2:NUMZ←(U+0.5)-NM
  NUMZ1←(U-0.5)-NM
  DEN←((MM×NN×(MM+NN+1))÷12)*0.5
  Z←NUMZ÷DEN
  Z1←NUMZ1÷DEN
  NUM←|(U-NM)
  DENC←((((NN+MM-1)×(DEN*2))÷(NN+MM-2))-(((NUM-0.5)*2)÷(NN+MM-2)))*0.5
  DENC1←((((NN+MM-1)×(DEN*2))÷(NN+MM-2))-(((NUM+0.5)*2)÷(NN+MM-2)))*0.5
  TC←(NUM-0.5)÷DENC
  TC1←(NUM+0.5)÷DENC1
  →(U≤NM)/SECOND
  PVI←((NORMCDF Z)+((NN+MM-2) TDISTN TC))÷2
  PV←((1-(NORMCDF Z1))+(1-((NN+MM-2) TDISTN TC1)))÷2
  →L4
 SECOND:PVI←((NORMCDF Z)+(1-((NN+MM-2) TDISTN TC)))÷2
  PV←((1-(NORMCDF Z1))+((NN+MM-2) TDISTN TC1))÷2
 L4:PV3←2×(⌊/(PV,PVI))
  →(PV3≤1)/N5
  PV3←1
 N5:PVM←(3,1)⍴(PVI,PV,PV3)
  'THE SUM OF THE X RANKS IS: ',(⍕RX),'.  THE U STATISTIC EQUALS: ',(⍕U1)
  ' '
 ⍝          LOGICAL STATEMENT FOR VARIANCE OUTPUT
  →(AA=2)/VAR
  ' THE P-VALUE FOR H0: M(X) = M(Y) VERSUS H1: M(X) ',(⍕LOGIC[D;1]),' M(Y)
   IS: ',(4⍕PVM[D;1])
  ' '
  →L8
 VAR:PVM←(3,1)⍴(PV,PVI,PV3)
  'THE P-VALUE FOR H0: V(X) = V(Y) VERSUS H1: V(X) ',(⍕LOGIC[D;1]),' V(Y)
   IS: ',(4⍕PVM[D;1])
  ' '
  →0
 L8:'WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE SHIFT IN LOCATION(M(Y) -
   M(X))?     (Y/N).'
  BB←⍞
  →(BB='N')/0
 L10:CC←INPUT 5
  ALPHA←(100-CC)÷200
 ⍝    ROUTE TO NORMAL APPROX. FOR CONF. INT. OF LARGER SAMPLE SIZES
  →((NM×2)>80)/L5
 ⍝           COMPUTING CONFIDENCE INTERVALS BY EXACT P-VALUE
  CDF←P
 ⍝      INDEX POSITION OF VALUE IN CDF ≤ ALPHA
  INDEX←(+/(CDF≤ALPHA))
  →(INDEX>0)/L6
  INDEX←1
  →L6
 ⍝         COMPUTING CONFIDENCE INTERVALS USING NORMAL APPROX. W/C.F.
 L5:UALPHA←(DENOMZ×(NORMPTH ALPHA))+NM-0.5
 ⍝          ROUND UALPHA DOWN AND INCREMENT BY ONE
  INDEX←⌊UALPHA+1
 L6:IPX←NN,INDEX
  CI←IPX CONFMW A
  ' A ',(⍕CC),' PERCENT CONFIDENCE INTERVAL FOR THE SHIFT IN '
  ' LOCATION BETWEEN POPULATIONS X AND Y IS:'
  ' '
  '                    ( ',(⍕CI[1]),' ≤ M(Y) - M(X) ≤ ',(⍕CI[2]),' )'
  ' '
  →0
 E1:'ERROR: THE SIZE OF YOUR SAMPLE IS LESS THAN TWO; TRY AGAIN.'
  ' '
  →B1
 E3:'ERROR: YOU HAVE ENTERED MORE THAN ONE VALUE. TRY AGAIN.'
  ' '
  →B3
 E2:'ERROR: YOU HAVE NOT ENTERED A VALUE OF 1 OR 2; TRY AGAIN.'
  ' '
  →B2
 E4:'ERROR: YOU HAVE NOT ENTERED A VALUE OF 1, 2, OR 3; TRY AGAIN.'
  ' '
  →(AA=2)/N2
  →B6
 ∇


 ∇ NPSLR;N;SUMX;SUMY;XBAR;YBAR;SUMX2;SUMXY;B;A;WW;XX;BB;U;D;ALPHA;P;CC;CDF
   ;TALPHA;NN;CI;SLOPES;RR;SR;YY;DENOM;INDEX;FF;X;Y;Q;R;CHA;PV
 ⍝   PROGRAM CONDUCTS NONPARAMETRIC LINEAR REGRESSION. THE LEAST SQUARES
 ⍝ ESTIMATED REGRESSION LINE IS COMPUTED WITH HYPOTHESIS TESTING AND
 ⍝ CONFIDENCE INTERVAL AVAILABLE FOR THE SLOPE B. IF B DOES NOT LIE IN
 ⍝ THE C.I. AN ALTERNATE REGRESSION LINE IS PROPOSED. SUBPROGRAMS CALLED
 ⍝ ARE: SPMANP, KENDALP, NORMPTH, INPUT, AND CONFLR.
  ⎕PP←5
 ⍝                      INPUT DATA
  R←INPUT 2
  Q←1↑R
  X←1↓(Q+1)↑R
  Y←(Q+1)↓R
 ⍝                ASSIGN THE SIZE OF X (AND Y) TO N
  N←⍴X
 ⍝                  COMPUTE THE SUM OF X'S AND Y'S
  SUMX←+/X
  SUMY←+/Y
 ⍝                  COMPUTE THE MEAN OF X AND Y
  XBAR←SUMX÷N
  YBAR←SUMY÷N
 ⍝              COMPUTE THE SUM OF THE X'S SQUARED
  SUMX2←+/(X*2)
 ⍝                  COMPUTE THE SUM OF X TIMES Y
  SUMXY←+/(X×Y)
 ⍝  COMPUTE 'B', THE SLOPE OF THE ESTIMATED LEAST SQUARES REGRESSION LINE
  B←((N×SUMXY)-(SUMX×SUMY))÷((N×SUMX2)-(SUMX*2))
 ⍝                COMPUTE 'A', THE Y-INTERCEPT
  A←YBAR-(B×XBAR)
  FF←'N'
  ' THE LEAST SQUARES ESTIMATED REGRESSION EQUATION IS: '
  ' '
  '                 Y = ',(⍕A),' + ',(⍕B),'X.'
  ' '
  'DO YOU WISH TO ENTER SOME X VALUES TO GET THE PREDICTED Y''S? (Y/N).'
  WW←⍞
  →(WW='N')/L1
 L2:'ENTER X VALUES.'
  XX←⎕
 ⍝                   CALCULATE PREDICTED Y'S
  YY←A+B×XX
  'THE PREDICTED Y VALUES ARE: ',(⍕YY)
  ' '
  ' WOULD YOU LIKE TO RUN SOME MORE X VALUES? (Y/N).'
  WW←⍞
  →(WW='Y')/L2
 L1:'WOULD YOU LIKE TO TEST HYPOTHESIS ON B, THE SLOPE OF THE EQUATION? (Y/
  N).'
  WW←⍞
  →(WW='N')/L3
  'ENTER THE HYPOTHESIZED SLOPE.'
  BB←⎕
 ⍝               COMPUTE UI'S
  U←Y-(BB×X)
  CHA←'> '
 ⍝  CALL SPMANP TO COMPUTE RHO AND ASSOCIATED P-VALUES.
  D←X SPMANP U
  →(D[1]>0)/L11
  CHA←'< '
 L11:PV←2×D[2]
```

```
  →(PV≤1)/L19
  PV←1
 L19:' SPEARMAN''S R EQUALS: ',(4⍕D[1])
  ' '
  'THE P-VALUE FOR H0: B = ',(⍕BB),' VERSUS H1: B ',(⍕CHA[1 2]),(⍕BB),' IS:
  ',(4⍕D[2])
  ' '
  'THE P-VALUE FOR THE TWO SIDED TEST OF HYPOTHESIS IS: ',(4⍕PV)
  ' '
 ⍝    IF USING THE NEW REGRESSION EQUATION BASED ON MEDIANS, EXIT HERE.
 L3:→(FF='Y')/0
 ⍝             COMPUTE CONFIDENCE INTERVALS ON B
  'WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE SLOPE? (Y/N).'
  WW←⍞
  →(WW='N')/0
 L10:CC←INPUT 5
 ⍝      CHANGE ENTERED VALUE TO ALPHA
  ALPHA←(100-CC)÷200
 ⍝    ROUTE TO NORMAL APPROX. FOR CONF. INT. OF LARGER SAMPLE SIZES
  →(N>12)/L5
 ⍝          COMPUTING CONFIDENCE INTERVALS BY EXACT P-VALUE
  P←KENDALP N
  CDF←P[2;]
 ⍝       INDEX POSITION OF VALUE IN CDF ≤ ALPHA
  INDEX←(+/(CDF≤ALPHA))
  →(INDEX>0)/L6
  INDEX←1
  →L6
 ⍝           COMPUTING CONFIDENCE INTERVALS USING NORMAL APPROX. W/C.F.
 L5:DENOM←((N×(N-1)×((2×N)+5))÷18)*0.5
  TALPHA←DENOM×(|(NORMPTH ALPHA))
  →L9
 L6:TALPHA←P[3;INDEX]
 L9:CI←X CONFLR Y
  NN←1↑CI
  SLOPES←1↓CI
  RR←⌊((NN-TALPHA)÷2)
  →(RR=0)/L20
  RR←1
 L20:SR←⌈(1+((NN+TALPHA)÷2))
  →(SR≤(⍴SLOPES))/L21
  SR←⍴SLOPES
 L21:' A ',(⍕CC),' PERCENT CONFIDENCE INTERVAL FOR B, THE SLOPE OF '
   ' THE ESTIMATED REGRESSION LINE, IS:'
   ' '
   '                      ( ',(⍕SLOPES[RR]),' < B < ',(⍕SLOPES[SR]),' ).'
   ' '
  ⍝ IF B OUTSIDE THE C.I. CALCULATE NEW EQUATION BASED ON MEDIANS
  ⍝→((B≥SLOPES[RR])∧(B≤SLOPES[SR]))/0
  ⍝                     ORDER X AND Y
   X←X[⍋X]
   Y←Y[⍋Y]
  ⍝    CHECK TO SEE IF THE SIZE OF SS IS EVEN OR ODD FOR FINDING MEDIANS
   →((2|NN)=0)/S1
  ⍝              COMPUTE MEDIAN FOR ODD CASE
   B←SLOPES[((NN+1)÷2)]
   →S2
  ⍝              COMPUTE MEDIAN FOR EVEN CASE
  S1:B←(SLOPES[(NN÷2)]+SLOPES[((NN÷2)+2)])÷2
  ⍝      DO THE SAME FOR THE X AND Y VECTORS
  S2:→((2|N)=0)/S3
   YBAR←Y[((N+1)÷2)]
   XBAR←X[((N+1)÷2)]
   →OUT
  S3:YBAR←(Y[(N÷2)]+Y[((N÷2)+2)])÷2
   XBAR←(X[(N÷2)]+X[((N÷2)+2)])÷2
  ⍝               COMPUTE NEW INTERCEPT 'A'
  OUT:A←YBAR-(B×XBAR)
   'THE LEAST SQUARES ESTIMATOR OF B LIES OUTSIDE THE CONFIDENCE INTERVAL.'
   'DISCARD THE LEAST SQUARES EQUATION AND USE:'
   ' '
   '                 Y = ',(⍕A),' + ',(⍕B),'X '
   ' '
   '    THIS EQUATION IS BASED ON THE MEDIANS OF THE X AND Y DATA AND '
   'THE MEDIAN OF THE TWO-POINT SLOPES CALCULATED FOR THE CONFIDENCE INTERV
   AL ON B.'
   ' '
  ⍝    ALLOW USER TO DO SOME ANALISIS ON NEW EQUATION
   'DO YOU WISH TO ENTER SOME X VALUES TO GET PREDICTED Y''S FROM THE NEW E
   QUATION?   (Y/N).'
   FF←⍞
   →(FF='Y')/L2
   FF←'Y'
   →L1
  ∇
 ∇ SIGN;A;C;B;D;PVAL;X;MO;N;CDF;ALPHA;CI;Y;AA;BB;CC;DD;PV;PVI;NNN;KPOS;ORD
   D;YY;ORDX;KALPHA;Z;Z1;QUA;WW;PVM;PV3;R;Q
  ⍝      THIS FUNCTION USES THE ORDINARY SIGN TEST TO CALCULATE THE K
  ⍝ STATISTIC, P-VALUE, AND CONFIDENCE INTERVAL AS A TEST FOR MEDIANS.
  ⍝ THE LAST OPTION WILL DISPLAY A TABLE OF CONFIDENCE INTERVALS OF
  ⍝ ORDERED STATISTICS WITH CONFIDENCE COEFFICIENTS.
  ⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: BINOM, NORMCDF, NORMPTH,
  ⍝ INPUT,AND QUANC.
   ' '
   '       DID YOU ENTER THIS PROGRAM FOR THE SOLE PURPOSE OF GENERATING CON
  FIDENCE INTERVALS FOR A SPECIFIED SAMPLE SIZE AND QUANTILE? (Y/N).'
   WW←⍞
   →(WW='Y')/B4
   '       THE NULL HYPOTHESIS STATES - THE POPULATION MEDIAN (M) IS EQUAL T
  O THE    HYPOTHESIZED MEDIAN (MO); H0: M = MO. '
   ' '
   '            WHICH ALTERNATIVE DO YOU WISH TO TEST?'
   ' '
 B3:' ENTER:   1 FOR H1: M < MO;   2 FOR H1: M > MO;    3 FOR H1: M ≠ MO.'
   C←⎕
   →((C≠1)∧(C≠2)∧(C≠3))/E3
 B2:'ENTER:  1 FOR SINGLE-SAMPLE PROBLEM;   2 FOR PAIRED-SAMPLE PROBLEM.'
   AA←⎕
   →((AA≠1)∧(AA≠2))/E2
   →(AA=2)/L9
  ⍝         INPUT DATA FOR SINGLE SAMPLE CASE
   X←INPUT 1
   NNN←⍴X
   MO←INPUT 3
   D←X-MO
   →L11
  ⍝                          PAIRED SAMPLE CASE
 L9:R←INPUT 2
   Q←1↑R
   X←1↓(Q+1)↑R
   Y←(Q+1)↓R
   DD←X-Y
   NNN←⍴DD
   MO←INPUT 4
   D←(X-Y)-MO
  ⍝                          COMPRESS D TO REMOVE ZEROS
 L11:A←(D≠0)/D
  ⍝                      RECORD LENGTH OF A AND ASSIGN TO N
   N←⍴A
  ⍝                         KEEPING TRACK OF POSITIVE SIGNS
   KPOS←+/(A>0)
   →(N≥30)/NORM
   PVAL←BINOM N
   →(KPOS>0)/P1
   PVI←1
   →P2
 P1:PVI←1-PVAL[KPOS]
 P2:PV←PVAL[(KPOS+1)]
   →L6
  ⍝ IF N IS GREATER THAN 30 USE NORMAL APPROX W/ CONTINUITY CORRECTION
 NORM:Z1←((KPOS-0.5)-(0.5×N))÷(0.5×(N*0.5))
   Z←((KPOS+0.5)-(0.5×N))÷(0.5×(N*0.5))
   PV←NORMCDF Z
   PVI←1-(NORMCDF Z1)
  ⍝         IF PAIRED SAMPLE TEST GO TO L17 FOR OUTPUT STATEMENT
 L6:PV3←2×(⌊/(PV,PVI))
   →(PV3≤1)/N5
   PV3←1
 N5:PVM←(3,1)⍴(PV,PVI,PV3)
   'COMPUTATIONS ARE BASED ON A SAMPLE SIZE OF: ',(⍕N)
   ' '
   'THE TOTAL NUMBER OF POSITIVE SIGNS IS: ',(⍕KPOS)
   ' '
   →(AA=2)/L17
   'THE P-VALUE FOR H0: M = ',(⍕MO),' VERSUS H1: M ',(⍕LOGIC[C;1]),' ',(⍕MO)
  ,' IS:',(4⍕PVM[C;1])
   ' '
   →L18
 L17:'THE P-VALUE FOR COMPARING THE MEDIAN OF THE POPULATION OF '
   'DIFFERENCES TO THE HYPOTHESIZED MEDIAN,'
   ' '
   'H0: M(X-Y) = ',(⍕MO),' VERSUS H1: M(X-Y) ',(⍕LOGIC[C;1]),' ',(⍕MO),', IS
  :',(4⍕PVM[C;1])
   ' '
 L18:'WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE MEDIAN? (Y/N).'
   BB←⍞
   →(BB='Y')/L16
   →QUANT
  ⍝                INPUT SIZE OF CONFIDENCE INTERVAL
 L16:CC←INPUT 5
   ALPHA←(100-CC)÷200
   →(NNN≥30)/NORM1
  ⍝         COMPUTING CONFIDENCE INTERVALS BY EXACT P-VALUE
   CDF←BINOM NNN
  ⍝                INDEX POSITION OF CDF FOR ALPHA ÷ 2
   B←+/(CDF≤ALPHA)
   →(B>0)/SKIP
 B←1
 →SKIP
⍝ COMPUTING CONFIDENCE INTERVALS BY NORMAL APPROX.
NORM1:KALPHA←((0.5×(NNN*0.5))×(NORMPTH ALPHA))+(0.5×NNN)-0.5
⍝  ROUND KALPHA DOWN TO NEAREST INTEGER AND INCREMENT BY ONE
 B←⌊KALPHA+1
⍝                      IF SINGLE SAMPLE CASE GO TO L7
SKIP:→(AA=1)/L7
⍝           CALCULATE AND PRINT OUT CONF. INT. FOR PAIRED SAMPLE CASE
L5:ORDD←DD[⍋DD]
 YY←⍴ORDD
 CI←ORDD[B],ORDD[(YY-(B-1))]
 'A ',(⍕CC),' PERCENT CONFIDENCE INTERVAL FOR THE MEDIAN OF THE '
 'POPULATION OF DIFFERENCES IS:'
 ' '
 '                     ( ',(⍕CI[1]),' ≤ MEDIAN(X-Y) ≤ ',(⍕CI[2]),' )'
 ' '
 →QUANT
⍝                     CALCULATE AND PRINT OUT CONF. INT. FOR ONE SAMPLE CASE
L7:ORDX←X[⍋X]
 YY←⍴ORDX
 CI←ORDX[B],ORDX[(YY-(B-1))]
 'A ',(⍕CC),' PERCENT CONFIDENCE INTERVAL FOR THE MEDIAN OF THE POPULATIO
 N IS: '
 ' '
 '              ( ',(⍕CI[1]),' ≤ MEDIAN ≤ ',(⍕CI[2]),' )'
 ' '
QUANT:'WOULD YOU LIKE CONFIDENCE INTERVALS FOR A SPECIFIED QUANTILE? (Y/N
 ).'
 WW←⎕
 →(WW='Y')/B1
 →0
B4:'ENTER DESIRED SAMPLE SIZE (SINGLE INTEGER VALUE).'
 NNN←⎕
 →((⍴NNN)>1)/E4
 →((1|NNN)≠0)/E4
B1:'ENTER DESIRED QUANTILE; FOR EXAMPLE: ENTER 20, FOR THE 20TH QUANTILE.
 '
 QUA←⎕
 →((QUA≤0)∨(QUA>100))/E1
 NNN QUANC QUA
 ' '
 ' ***** THIS TABLE GIVES CONFIDENCE COEFFICIENTS FOR VARIOUS INTERVALS W
 ITH '
 'ORDER STATISTICS AS THE END POINTS FOR THE ',(⍕QUA),'TH QUANTILE.'
 ' '
 →0
E1:'ERROR: THE QUANTILE VALUE MUST LIE BETWEEN 0 AND 100; TRY AGAIN.'
 ' '
 →B1
E2:'ERROR: YOU HAVE NOT ENTERED A VALUE OF 1 OR 2; TRY AGAIN.'
 ' '
 →B2
E3:'ERROR: YOU HAVE NOT ENTERED A VALUE OF 1, 2, OR 3; TRY AGAIN.'
 ' '
 →B3
E4:'ERROR: YOU HAVE NOT ENTERED A SINGLE, INTEGER VALUE; TRY AGAIN.'
 ' '
 →B4
 ∇


 ∇ SPEARMAN;X;Y;A;Q;R;CHA;PV
⍝       THIS FUNCTION COMPUTES THE SPEARMAN R STATISTIC WHICH MEASURES
⍝ THE DEGREE OF CORRESPONDENCE BETWEEN RANKINGS OF TWO SAMPLES. THE P-
⍝ VALUE IS GIVEN FOR TESTING ONE AND TWO-SIDED HYPOTHESIS OF ASSOC.
⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: TIES, TIESK, SPEARP,
⍝ INPUT, SPAPROX, INTERP, AND THE VARIABLE PMATSP.
 R←INPUT 2
 Q←1↑R
 X←1↓(Q+1)↑R
 Y←(Q+1)↓R
⍝ CALL SPMANP TO CALCULATE THE STATISTIC AND ASSOCIATED P-VALUES
 A←X SPMANP Y
 →(A[1]>0)/L1
 CHA←'INDIRECT'
 →L2
L1:CHA←'DIRECT'
L2:PV←2×A[2]
 →(PV≤1)/L3
 PV←1
L3:'SPEARMAN''S R EQUALS ',(4⍕A[1])
 ' '
 'THE P-VALUE FOR H0: NO ASSOCIATION EXISTS VERSUS'
 '                H1: ',(⍕CHA),' ASSOCIATION EXISTS IS:',(4⍕A[2])
 ' '
 'THE P-VALUE FOR THE TWO-SIDED TEST OF HYPOTHESIS IS:',(4⍕PV)
 ' '
 ∇
```

```
    ∇ WILCOX;A;B;D;E;F;PV2;Z1;Z;DEN;NUMZ;NUMZ1;PVAL;X;MO;N;TPLUS;CDF;TALPHA;A
      LPHA;H;CI;Y;AA;BB;CC;NN;DD;PV;POS;TPOS;NM;PVI;TPOS1;NNN;C;PVM;PV3;R;Q;T
      RAP;NUM;DENT;DENT1;TC;TC1
   ⍝ THIS PROGRAM USES THE WILCOXON SIGNED RANK TEST TO CALCULATE THE  T
   ⍝ STATISTIC, P-VALUE, AND CONFIDENCE INTERVAL AS A TEST FOR MEDIANS.
   ⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: TIES, WILP, NORMCDF,
   ⍝ NORMPTH, CONFW, AND INPUT.
    ' '
    '        THE NULL HYPOTHESIS STATES - THE POPULATION MEDIAN (M) IS EQUAL T
   O THE   HYPOTHESIZED MEDIAN (MO); H0: M = MO.'
    ' '
    '                WHICH ALTERNATIVE DO YOU WISH TO TEST?'
    ' '
  B3:' ENTER:    1 FOR H1: M < MO;    2 FOR H1: M > MO;     3 FOR H1: M ≠ MO.'
    C←⎕
    →((C≠1)∧(C≠2)∧(C≠3))/E3
  B2:'ENTER:    1 FOR SINGLE-SAMPLE PROBLEM;    2 FOR PAIRED-SAMPLE PROBLEM.'
    AA←⎕
    →((AA≠1)∧(AA≠2))/E2
    →(AA=2)/L9
   ⍝                      INPUT DATA FOR SINGLE SAMPLE CASE
    X←INPUT 1
    NNN←ρX
    MO←INPUT 3
    D←X-MO
    →L11
   ⍝                      PAIRED SAMPLE CASE
  L9:R←INPUT 2
    Q←1↑R
    X←1↓(Q+1)↑R
    Y←(Q+1)↓R
    DD←X-Y
    NNN←ρDD
    MO←INPUT 4
    D←(X-Y)-MO
   ⍝                      COMPRESS D TO REMOVE ZEROS
  L11:A←(D≠0)/D
   ⍝                   RECORD LENGTH OF A AND ASSIGN TO N
    N←ρA
   ⍝                     KEEPING TRACK OF POSITIVE SIGNS
    POS←(A>0)
   ⍝              TAKE THE ABSOLUTE VALUE OF A; ASSIGN TO B AND ORDER B
    B←|A
    B←B[⍋B]
   ⍝       REORDER POSITIVE SIGNS TO COINCIDE WITH PROPER POSITIONS IN B
    POS←POS[⍋(|A)]
   ⍝                      CALL FUNCTION TO BREAK TIES
    E←1 TIES B
   ⍝       CALCULATE TPOS BY ADDING ACROSS ALL POSITIVE VALUES OF E
    TPOS←+/(POS×E)
    TPOS1←TPOS
   ⍝          GIVES SIZE OF LEFT TAIL OF PROBABILITY DISTRIBUTION
    NM←(⌊((+/ιN)÷2))+1
   ⍝              GO TO STATEMENTS BASED ON LENGTH OF VECTOR E
    →(N>16)/L3
   ⍝                   GENERATE NULL DISTRIBUTION FOR TPLUS
    F←WILP N
   ⍝  IF TPOS FALLS IN LEFT HALF OF PROB DIST CALCULATE PVALUE AS NORMAL
    →(TPOS≤(NM-1))/TREG
   ⍝           OTHERWISE USE THE NEGATIVE T STATISTIC
    TPOS←(+/ιN)-TPOS
   ⍝ IF TPOS IS FRACTIONAL USE BOTH THE INTEGER ABOVE AND BELOW AS TPLUS
  TREG:→((1|TPOS)=0)/NON
    TPLUS←⌊TPOS
  P1:PV←1-((F[TPLUS]+F[TPLUS+1])÷2)
    PVI←(F[TPLUS+1]+F[TPLUS+2])÷2
    →CHECK
  NON:→(TPOS>0)/GO
    PV←1
    →P2
  GO:PV←1-F[(TPOS)]
  P2:PVI←F[(TPOS+1)]
  CHECK:→(TPOS1≤(NM-1))/L6
    PV2←PV
    PV←PVI
    PVI←PV2
    →L6
   ⍝  COMPUTE NORMAL APPROX. W/CONTINUITY CORRECTION FACTOR
  L3:TRAP←(N×(N+1))÷4
    NUMZ←(TPOS+0.5)-TRAP
    NUMZ1←(TPOS-0.5)-TRAP
    DEN←((N×(N+1)×((2×N)+1))÷24)*0.5
    Z←NUMZ÷DEN
    Z1←NUMZ1÷DEN
   ⍝ COMPUTE STUDENT T APPROXIMATION WITH CONTINUITY CORRECTION FACTOR
    NUM←|(TPOS-TRAP)
    DENT←(((N×(DEN*2))÷(N-1))-(((NUM-0.5)*2)÷(N-1)))*0.5
    DENT1←(((N×(DEN*2))÷(N-1))-(((NUM+0.5)*2)÷(N-1)))*0.5
    TC←(NUM-0.5)÷DENT
    TC1←(NUM+0.5)÷DENT1
   ⍝ COMPUTE AVERAGE OF TC AND ZC
  →(TPOS≤((+/⍳N)÷2))/SECOND
  PV←((1-(NORMCDF Z1))+(1-((N-1) TDISTN TC1)))÷2
  PVI←((NORMCDF Z)+((N-1) TDISTN TC))÷2
  →L6
 SECOND:PV←((1-(NORMCDF Z1))+((N-1) TDISTN TC1))÷2
  PVI←((NORMCDF Z)+(1-((N-1) TDISTN TC)))÷2
 L6:PV3←2×(⌊/(PV,PVI))
  →(PV3≤1)/N5
  PV3←1
 N5:PVM←(3,1)⍴(PVI,PV,PV3)
  'COMPUTATIONS ARE BASED ON A SAMPLE SIZE OF: ',(⍕N)
  ' '
  'THE TOTAL SUM OF POSITIVE RANKS IS: ',(⍕TPOS1)
  ' '
 ⍝          IF PAIRED SAMPLE TEST GO TO L17 FOR OUTPUT STATEMENT
  →(AA=2)/L17
  'THE P-VALUE FOR H0: M = ',(⍕MO),' VERSUS H1: M ',(⍕LOGIC[C;1]),' ',(⍕MO
  ),' IS:',(4⍕PVM[C;1])
  ' '
  →L18
 L17:'THE P-VALUE FOR COMPARING THE MEDIAN OF THE POPULATION OF '
  'DIFFERENCES TO THE HYPOTHESIZED MEDIAN,'
  ' '
  'H0: M(X-Y) = ',(⍕MO),' VERSUS H1: M(X-Y) ',(⍕LOGIC[C;1]),' ',(⍕MO),', I
  S:',(4⍕PVM[C;1])
  ' '
 L18:'WOULD YOU LIKE A CONFIDENCE INTERVAL FOR THE MEDIAN? (Y/N).'
  BB←⍞
  →(BB='Y')/L16
  →0
 L16:CC←INPUT 5
  ALPHA←(100-CC)÷200
 ⍝         ROUTE TO NORMAL APPROX. FOR CONF INT OF LARGE SAMPLE SIZE
  →(NNN>15)/L4
 ⍝         COMPUTING CONFIDENCE INTERVALS BY EXACT P-VALUE
  CDF←WILP NNN
 ⍝                 INDEX POSITION OF CDF FOR ALPHA ÷ 2
  TALPHA←(+/(CDF≤ALPHA))
  →(TALPHA>0)/JUMP
  TALPHA←1
  →JUMP
 ⍝              COMPUTING CONFIDENCE INTERVALS BY NORMAL APPROX. W/C.F.
 L4:DENOMZ←((2×NNN×(NNN+1)×((2×NNN)+1))÷3)*0.5
  TALPHA←(((DENOMZ×(NORMPTH ALPHA))+(NNN×(NNN+1))-2)÷4)
 ⍝         ROUND TALPHA DOWN TO INTEGER VALUE AND INCREMENT BY ONE
  TALPHA←⌊TALPHA+1
 ⍝                 IF ONE SAMPLE CASE GO TO L7
 JUMP:→(AA=1)/L7
 ⍝               CALCULATE AND PRINT OUT CONF. INT. FOR PAIRED SAMPLE CASE
 L5:CI←TALPHA CONFW DD
  'A ',(⍕CC),' PERCENT CONFIDENCE INTERVAL FOR THE MEDIAN OF THE '
  'POPULATION OF DIFFERENCES IS:'
  ' '
  '                ( ',(⍕CI[1]),' ≤ MEDIAN(X-Y) ≤ ',(⍕CI[2]),' )'
  ' '
  →0
 ⍝               CALCULATE AND PRINT OUT CONF. INT. FOR ONE SAMPLE CASE
 L7:CI←TALPHA CONFW X
  'A ',(⍕CC),' PERCENT CONFIDENCE INTERVAL FOR THE MEDIAN OF THE POPULATIO
  N IS:'
  ' '
  '                 ( ',(⍕CI[1]),' ≤ MEDIAN ≤ ',(⍕CI[2]),' )'
  ' '
  →0
 E2:'ERROR: YOU HAVE NOT ENTERED A VALUE OF 1 OR 2; TRY AGAIN.'
  ' '
  →B2
 E3:'ERROR: YOU HAVE NOT ENTERED A VALUE OF 1, 2, OR 3; TRY AGAIN.'
  ' '
  →B3
 ∇
```

APPENDIX F

LISTINGS OF SUBPROGRAMS BASIC TO BOTH WORKSPACES

```
      ∇ CBN←BINOM N;N;P;X;K;CDF
     ⍝ THIS FUNCTION IS A SUBPROGRAM OF THE SIGN TEST (SIGN).  IT CALCULATES
     ⍝ THE CDF OF THE BINOMAL WHEN PROBABILITY = .5. N=SAMPLE SIZE.
      P←0.5
      K←0,⍳N
      CBN←+\(K!N)×(P*K)×((1-P)*N-K)
      ∇
```

```
      ∇ U←L CBIN R

     ⍝   THIS FUNCTION IS A SUBPROGRAM OF CONFIDENCE INTERVAL GENERATOR FOR
     ⍝ THE QTH QUANTILE.  IT RETURNS THE VALUE OF THE BINOMIAL CDF AT R,
     ⍝ WITH N,P=L WHERE N=SAMPLE SIZE AND P=PROBABILITY.

      U←(+\((¯1+⍳1+L[1])!L[1])×(L[2]*¯1+⍳1+L[1])×((1-L[2])*L[1]-¯1+⍳1+L[1]))[R+
      1]
      ∇
```

```
      ∇ CON←XX CONFLR Y1;BB;SS;AA:A;XR:YR;B:C;S
     ⍝       THIS FUNCTION IS A SUBPROGRAM OF NONPARAMETRIC LINEAR REGRESSION
     ⍝ (NPLR).  IT CALCULATES THE TWO-POINT SLOPE FOR EACH PAIR OF POINTS
     ⍝ (XI,YI) AND (XJ,YJ,) ALL I<J AND XI=XJ. ALL SLOPES ARE ORDERED AND
     ⍝ USED TO FIND THE CONFIDENCE INTERVAL FOR B, THE SLOPE OF ESTIMATED
     ⍝ EQUATION.  XX= X DATA SAMPLE AND Y1= Y DATA SAMPLE.

     ⍝ RECORD THE SIZE OF XX AND INITIALIZE VARIABLES
      BB←⍴XX
      SS←⍴0
      AA←0
     ⍝ THIS LOOP COMPRESSES XX AND YY DOWN TO WHERE THE XX < ALL OTHER XX'S
     L2:AA←AA+1
      A←(XX[AA]<XX)
      XR←A/XX
      YR←A/Y1
      B←⍴XR
      →(B=0)/L3
      C←0
     ⍝ THIS LOOP CALCULATES THE SLOPE OF EACH PAIR OF PAIRED DATA.
     L1:C←C+1
      S←(Y1[AA]-YR[C])÷(XX[AA]-XR[C])
      SS←SS,S
      →(C<B)/L1
     L3:→(AA<BB)/L2
      CON←(⍴SS),(SS[⍋SS])
      ∇
```

```
      ∇ CONFM←AA CONFMW BB;A;B;C;D:E;F;G;H
     ⍝       THIS FUNCTION IS A SUBPROGRAM OF THE MANN-WHITNEY TEST (MANW).
     ⍝ IT COMPUTES CONFIDENCE INTERVAL ENPOINTS FOR THETA, THE SHIFT IN
     ⍝ LOCATION BETWEEN X AND Y. AA=INDEX POSITION OF C.I. ENPOINT.
     ⍝ BB=COMBINED DATA SAMPLES.
     ⍝     ASSIGN SIZE OF X VECTOR TO A; INDEX POSITION FOR CONF INT TO B.
      A←AA[1]
      B←AA[2]
     ⍝            ASSIGN X VECTOR TO C; Y VECTOR TO D
      C←A↑BB
      D←A↓BB
     ⍝            REORDER X AND Y VECTOR VALUES TO ASCENDING ORDER
      C←C[⍋C]
      D←D[⍋D]
     ⍝            INITIALIZE VECTOR H AND VARIABLE F
      H←⍴0
      F←1
     ⍝INNER AND OUTER LOOPS CALCULATE ALL POSSIBLE DIFFERENCES; EVERY Y
     ⍝ ELEMENT MINUS EVERY X ELEMENT. H VECTOR STORES THESE DIFFERENCES.
     L2:G←1
     L1:E←D[F]-C[G]
      H←H,E
```

```
[22]   G←G+1
[23]   →(G≤(⍴C))/L1
[24]   F←F+1
[25]   →(F≤(⍴D))/L2
[26]  ⍝         REORDER H VECTOR VALUES TO ASCENDING ORDER
[27]   H←H[⍋H]
[28]  ⍝         INDEX APPROPRIATE CONF INT VALUES FROM H
[29]   CONFM←H[B],H[((⍴H)-(B-1))]
       ∇


       ∇ CONF←A CONFW B;C;D;E;G;H;N;F;A;B
[1]   ⍝      THIS FUNCTION IS A SUBPROGRAM OF THE WILCOXON SIGNED RANK TEST
[2]   ⍝ (WISIG). IT PROVIDES CONFIDENCE INTERVAL END POINTS BASED ON THE
[3]   ⍝ AVERAGES OF ALL PAIRS OF DATA SUCH THAT ALL XI ≤ XK. A= INDEX
[4]   ⍝ POSITION OF C. I. END POINT AND B= INPUT DATA.
[5]    C←B[⍋B]
[6]   ⍝         START H ACCUMULATION VECTOR OFF WITH ORGINAL VECTOR  VALUES
[7]    H←⍳0
[8]    H←H,C
[9]    D←0
[10]  ⍝  OUTSIDE LOOP INCREMENTS D AND RESETS E TO D
[11]  L2:D←D+1
[12]   E←D
[13]  ⍝  INSIDE LOOP GENERATES NEXT SET OF AVERAGES AND CONCATENATES TO ORIGINAL
        VECTOR
[14]  L1:E←E+1
[15]   F←(C[D]+C[E])÷2
[16]   H←H,F
[17]  ⍝            CONTINUE INNER LOOP UNTIL E EQUALS THE SIZE OF C
[18]   →(E<(⍴C))/L1
[19]  ⍝        CONTINUE OUTER LOOP UNTIL D EQUALS THE SIZE OF C LESS ONE
[20]   →(D<((⍴C)-1))/L2
[21]  ⍝               ORDER FINAL ACCUMULATED VECTOR H
[22]   H←H[⍋H]
[23]  ⍝                INDEX CONF INT VALUES OUT OF H
[24]   CONF←H[A],H[((⍴H)-(A-1))]
       ∇


       ∇ P←DF FDISTN X;A;M;N;RM;RN;LM;LN;SM;SN;M2;N2
[1]   ⍝     THIS FUNCTION IS SUBPROGRAM OF THE KRUSKAL-WALLIS TEST (KRWL) AND
[2]   ⍝ THE STUDENT T DIST.(TDISTN). IT CALCULATES APPROX. CUMULATIVE PROBS.
[3]   ⍝ OF X USING THE F DISTRIBUTION W/ DF=(M,N) DEGREES OF FREEDOM.
[4]
[5]   ⍝   TAKES VECTOR ARG.
[6]    A←,M×X÷N+X×M←1↑DF,N←1↓DF
[7]    →L×⍳(M>2)∨N>2
[8]   ⍝ TREAT THE 1ST. FOUR SPECIAL CASES.
[9]    →⍎'L',¯1 0 ⍕M,N
[10]  L11:P←(¯1○A*0.5)÷○÷2
[11]   →0
[12]  L12:P←A*0.5
[13]   →0
[14]  L21:P←1-(1-A)*0.5
[15]   →0
[16]  L22:P←A
[17]   →0
[18]  ⍝ BEGIN THE GENERAL CASES. INITIALIZE THE QUANTITIES.
[19]  L:RM←(1|M÷2),0/RN←1|N÷2
[20]   SM←(RM=0)×RM≤LM←⌊M2←0.5×M-2
[21]   SN←(RN=0)×RN≤LN←⌊N2←0.5×N-2
[22]  ⍝ TREAT SN,CASE N-EVEN FIRST.
[23]   →NODD×⍳1=2|N
[24]   →MEVEN×⍳0=N2
[25]   SN←1++/×\(1-A)∘.×1+(¯1+0.5×M)÷⍳N2
[26]   →MEVEN
[27]  NODD:→MEVEN×⍳1=N
[28]   SN←(1-A)*0.5
[29]  ⍝ TREAT THE PORTION OF SN THAT DOESN'T CHANGE W/ M ODD OR EVEN.
[30]   →(2+1↑⎕LC)×⍳0=LN
[31]   SN←SN×1++/×\(1-A)∘.×1+2×M2÷1+2×⍳LN
[32]  ⍝ TREAT THE M-EVEN SUBCASE
[33]   →((M=2),1=2|M)/MEVEN,3+1↑⎕LC
[34]   SN←SN××/1+÷2×⍳M2
[35]   →MEVEN
[36]  ⍝ TREAT THE M-ODD SUBCASE
[37]   SN←SN×2÷○1
[38]   →END×⍳1=M
[39]   SN←SN××/1+÷¯1+2×⍳0.5+M2
[40]  MEVEN:→MODD×⍳1=2|M
[41]   →END×⍳2=M
[42]   SM←1++/×\A∘.×1-÷2×⍳M2
[43]   →END
[44]  ⍝ NEXT TREAT THE SPECIAL CASES FOR ODD-M.
[45]  MODD:SM←(0×M=1)+((A*0.5)÷○÷2)×3=M
[46]   →END×⍳(1=M)∨3=M
[47]   SM←(A*0.5)×(÷○÷2)×1++/×\A∘.×1-÷1+2×⍳LM
[48]  END:P←(SN×A*M÷2)+(4×RN×RN-RM)+((RM×RN×¯1○A*0.5)÷○÷8)-2×RN×SM×(1-A)*0.5
       ∇
```

```
  ∇ XX←N INDEXPLS B;AA;BB;CC;DD;X1;NN;NC;N;B;C;D;E;F
 ⍝      THIS FUNCTION IS A SUBPROGRAM OF WRUSKAL-WALLIS (KRWL), AND MANN-
 ⍝ WHITNEY (MANW).  IT COMPUTES A MATRIX OF RANKS OF THE ORIGINAL DATA
 ⍝ WHERE THE FIRST ROW OF THE MATRIX IS THE FIRST SAMPLE.  IF TIES OCCUR
 ⍝ IN THE RANKING WITHIN A SAMPLE, SUCCEEDING TIES ARE INCREMENTED BY
 ⍝ ONE. N= SIZES OF THE SAMPLES BEING PASSED IN B. B= ALL DATA SAMPLES
 ⍝ COMBINED.

 ⍝          RECORD SIZE OF N TO DETERMINE NUMBER OF SAMPLES IN B
  AA←ρN
 ⍝                LOCATE LARGEST SAMPLE SIZE
  BB←⌈/N
 ⍝                ORDER B SMALLEST TO LARGEST
  DD←B[⍋B]
 ⍝         SET UP MATRIX OF SIZE REQUIRED TO STORE SAMPLE RANKING.
  XX←(AA,BB)ρ0
 ⍝         CONCATENATE 0 AND N
  NN←0,N
 ⍝       FIND CUMULATIVE SUMS OF SAMPLE SIZES.
  NC←+\NN
  CC←0
 ⍝ THIS LOOP INDEXES OUT ORGINAL SAMPLE VALUES FOR FURTHER CALCULATIONS.
 L1:CC←CC+1
 ⍝          LOCATE CC'TH ORIGINAL SAMPLE VALUES IN B
  X1←B[(NC[CC]+⍳N[CC])]
 ⍝          RECORD POSITIONS OF ELEMENTS OF X VECTOR IN B AND ASSIGN TO C
  C←DD⍳X1
  D←1
 ⍝ THIS LOOP DETECTS TIED INDEXED POSITIONS AND INCREMENTS THE INDEXING OF
  EACH SUCCESSIVE TIED POSITION BY ONE
 ⍝                TEST DTH ELEMENT OF C AGAINST REST OF C FOR TIES
 L2:E←(C[D]=D↓C)
 ⍝           SET F EQUAL TO THE APPROPRIATE SIZE (DTH SIZE) ZERO VECTOR
  F←Dρ0
 ⍝ CONCATENATE F AND VECTOR OF 0'S AND 1'S (1'S APPEAR WHEN TIE OCCURED)
  F←F,E
 ⍝                 ADD RESULTANT F VECTOR TO C
  C←C+F
  D←D+1
 ⍝                 CONTINUE FOR ENTIRE C VECTOR
  →(D≤(ρC))/L2
  XX[CC;]←C,(BB-N[CC])ρ0
  →(CC<AA)/L1
  ∇
```

```
  ∇ IN←INPUT A;B;X;Y
 ⍝ THIS FUNCTION DOES MOST OF THE INPUT PROMPTING AND ERROR CHECKING.
 ⍝ A=THE TYPE OF PROMPT DESIRED.
 ⍝ IT IS A SUBPROGRAM OF SIGN, WISIG, MANW, KEN, SPMAN, AND NPLR.
  →(L1,L2,L3,L4,L5)[A]
 L1:'ENTER THE DATA (MORE THAN TWO OBSERVATIONS ARE REQUIRED). '
  IN←⎕
  →((ρρIN)=0)/E1
  →((ρIN)=2)/E1
  →0
 L2:'ENTER X DATA (MORE THAN TWO OBSERVATIONS ARE REQUIRED).'
  X←⎕
  →((ρρX)=0)/E1
  →((ρX)=2)/E1
  'ENTER Y DATA (NUMBER OF Y ENTRIES MUST EQUAL NUMBER OF X ENTRIES).'
  Y←⎕
  →((ρρY)=0)/E2
  →((ρX)≠(ρY))/E2
  IN←(ρX),X,Y
  →0
 L3:'ENTER THE HYPOTHESIZED MEDIAN.'
  IN←⎕
  →((ρIN)>1)/E3
  →0
 L4:'ENTER THE HYPOTHESIZED MEDIAN FOR THE DIFFERENCES OF THE PAIRED DATA.'
  IN←⎕
  →((ρIN)>1)/E3
  →0
 L5:'             ENTER THE DESIRED CONFIDENCE COEFFICIENT; '
  ' '
  ' FOR EXAMPLE: ENTER 95, FOR A 95 PERCENT CONFIDENCE INTERVAL.'
  IN←⎕
  →((IN≤0)∨(IN>100))/E4
  →((1|IN)≠0)/E5
  →0
 E1:'ERROR: THE SIZE OF YOUR SAMPLE IS LESS THAN THREE; TRY AGAIN.'
  ' '
  →(A=1)/L1
  →L2
 E2:'ERROR: SAMPLE SIZES ARE NOT EQUAL; WANT TO TRY AGAIN? (Y/N).'
  B←⍞
  →(B='Y')/L2
  'ENTER RIGHT ARROW  →  TO QUIT.'
  B←⎕
 E3:'ERROR: THE HYPOTHESIZED MEDIAN MUST BE A SINGLE VALUE; TRY AGAIN.'
  ' '
  →(A=3)/L3
  →L4
 E4:'ERROR: THIS VALUE MUST LIE BETWEEN 0 AND 100; TRY AGAIN.'
  ' '
  →L5
 E5:'ERROR: THIS VALUE MUST BE AN INTEGER; TRY AGAIN.'
  ' '
  →L5
  ∇


  ∇ INTER←A INTERP B;C;D;E;F;G;FF;GG;PL
 ⍝      THIS FUNCTION IS A SUBPROGRAM OF THE KRUSKAL-WALLIS (KRWL),
 ⍝ KENDALL'S B (KEN), AND SPEARMAN'S R (SPMAN AND SPMAN1). WHEN
 ⍝ PASSED THE TEST STATISTIC AS THE LEFT ARGUMENT, IT CALCULATES THE
 ⍝ ASSOCIATED P-VALUE BY INTERPOLATION OF THE TABLE VALUES OF STATS.
 ⍝ AND CDF'S WHICH ARE PASSED AS THE RIGHT ARGUMENT.

 ⍝  SEPARATE THE CDF TABLE AND STATS. INTO SINGLE VARIABLES.
  FF←B[1;]
  GG←B[2;]
 ⍝    FIND WHERE A FIRST EXCEEDS OR EQUALS ONE OF THE TABLE VALUES.
  C←<\(A≥FF)
 ⍝   IF THE STAT. DOES NOT EQUAL ANY TABLE VALUES SET INTER = ¯1.
  →((+/C)=0)/L2
 ⍝       INDEX LOCATION OF FIRST OCCURENCE OF MATCH
  D←C⍳1
 ⍝         IF INDEXED POSITION EQUALS ONE INDEX P-VALUE OUT OF GG.
  →(D>1)/L1
  INTER←GG[D]
  →0
 ⍝         OTHERWISE CONDUCT INTERPOLATION TO GET PROPER P-VALUE.
 L1:E←FF[D-1]-FF[D]
  F←FF[D-1]-A
  G←GG[D]-GG[D-1]
  PL←(F×G)÷E
  INTER←GG[D-1]+PL
  →0
 L2:INTER←¯1
  ∇


  ∇ KENP←KENDALPP N;A;B;C;D;E;NN;X;XX;AA;F;T;G;P;TPL
 ⍝        THIS FUNCTION IS A SUBPROGRAM OF KENDALL'S B (KEN) AND NON-
 ⍝ PARAMETRIC LINEAR REGRESSION (NPLR).  IT CALCULATES THE CUMULATIVE
 ⍝ DISTRIBUTION OF B FOR A SAMPLE SIZE N.
  X← 1 1
 ⍝    INITIALIZE FREQUENCIES FOR X FOR SAMPLE SIZE NN.
  NN←2
 ⍝        DETERMINE SIZE OF RIGHT PROB. TAIL VECTOR
  G←(⌊((N×(N-1))÷4))+1
 ⍝ OUTER LOOP INCREMENTS THROUGH THE N SAMPLE SIZES TILL THE DESIRED ONE IS
   GENERATED
 L1:D←⍴0
  C←0
  F←⍴X
  NN←NN+1
  B←((NN×(NN-1))÷2)+1
  A←B
 ⍝ INNER LOOP GENERATES NN+1 FREQUENCIES FROM THE VECTOR OF NN FREQUENCIES
 L2:A←A-1
  C←C+X[(B-A)]
  D←D,C
  AA←0
 ⍝      IF SIZE OF D EQUALS NN AND INDEXES OF X STILL REMAIN GO TO L4
  →(((⍴D)=NN)∧((B-A)<F))/L4
 ⍝        OTHERWISE CONTINUE TO INCREMENT THRU OLD FREQS
  →((B-A)<F)/L2
 ⍝       WHEN LEFT HALF OF NN+1 VECTOR IS COMPLETE GO TO L5
  →L5
 ⍝       THIS LOOP ALLOWS ONLY NN TERMS TO BE USED TO GENERATE NEW VECTOR
 L4:A←A-1
  AA←AA+1
  C←C+X[(B-A)]-X[AA]
  D←D,C
  →((⌊((B-1)÷2)+1)>(B-A))/L4
 ⍝              INVERT VECTOR D AND ASSIGN TO E
 L5:E←⊖D
 ⍝  FOR NN OF APPROPRIATE SIZE EITHER DROP FIRST VALUE OFF E OR NOT
  →(NN=3,6,7,10,11,14,15,18,19)/L3
  E←1↓E
 ⍝  COMPLETE NEW VECTOR OF FREQS X BY CONCATENATING D WITH E
 L3:X←D,E
 ⍝           CONTINUE UNTIL SIZE OF SAMPLE N IS REACHED
  →(NN<N)/L1
 ⍝    GENERATE VECTOR OF CORRESPONDING P STATS OF PROPER SIZE G
```

```
[43]  P←0,⍳(G-1)
[44]  ⍝     CALCULATE B STATS FROM THE P VECTOR
[45]  T←|(((4×P)÷(N×(N-1)))-1)
[46]  ⍝    TAKE ONLY G ENTRIES FROM THE FREQUENCY VECTOR
[47]  TPL←T×(N×(N-1))÷2
[48]  XX←G↑X
[49]  ⍝    CHANGE FREQS TO CDF VALUES AND OUTPUT B STATS W/APPRO. CDF VALUES
[50]  KENP←(3,G)⍴T,((+\XX)÷(!N)),TPL
      ∇
```

```
      ∇ MAN←N MANWP M;F;Q;P;S;T;U;V;B;NN;N;M;UU;MM
[1]   ⍝       THIS FUNCTION IS A SUBPROGRAM OF THE MANN-WHITNEY TEST (MANW).
[2]   ⍝ IT GENERATES THE CUMULATIVE DIST. FOR THE U STATISTIC. N = SIZE OF
[3]   ⍝ LARGER SAMPLE; M = SIZE OF OTHER SAMPLE.
[4]
[5]   ⍝COMPUTE NUMBER OF TERMS TO BE INCLUDED IN LEFT TAIL DISTRIBUTION LESS 1
[6]    MM←(⌊((N×M)÷2))+1
[7]   ⍝           SET F VECTOR EQUAL TO 1 CONCATENATED WITH  MM ZERO'S
[8]    F←1,MM⍴0
[9]   ⍝            SET P EQUAL TO THE MINIMUM OF N+M OR MM
[10]   P←⌊/((N+M),MM)
[11]  ⍝            SET Q EQUAL TO THE MINIMUM OF M OR MM
[12]   Q←⌊/M,MM
[13]  ⍝    GO TO LINE DENOM IF MM IS LESS THAN N+1(SIZE OF X+1)
[14]   →DENOM×⍳(MM<N+1)
[15]  ⍝  IF MM≥N+1 GENERATE FIRST BLOCK OF RECURSIVE RESULTS USING NUM LOOP
[16]  NUM:T←N+1
[17]  L2:U←MM+1
[18]  ⍝ PRIMARY FORMULA USED IN GENERATION OF FIRST BLOCK OF RECURSIVE RESULTS
[19]  L1:F[U]←F[U]-F[U-T]
[20]  ⍝ ASSIGNS NEW DECREMENTED VALUE TO U AND TESTS IF T < THIS NEW U
[21]   →L1×⍳(T<U←U-1)
[22]   →L2×⍳(P≥T←T+1)
[23]  ⍝          GENERATE FINAL RECURSIVE RESULTS USING DENOM LOOP
[24]  DENOM:S←1
[25]  L4:V←S+1
[26]  ⍝    PRIMARY FORMULA USED IN GENERATION OF FINAL RECURSIVE RESULTS
[27]  L3:F[V]←F[V]+F[V-S]
[28]   →L3×⍳((MM+1)≥V←V+1)
[29]   →L4×⍳(Q≥S←S+1)
[30]  ⍝      CONVERT FREQUENCY TABLE TO CDF VALUES FOR FINAL OUTPUT
[31]   MAN←(+\F)÷(N!(N+M))
      ∇
```

```
      ∇ Z←NORMCDF X;A;B;C;D
[1]   ⍝ EVALUATES NORMAL CDF AT VECTOR X. FOR |X<4, 26.2.11 IN ABRAMOWITZ AND
[2]   ⍝ STEGUN, P. 932, IS USED. FOR LARGE X, THE CONTINUED FRACTION IN WALL,
[3]   ⍝P. 357, 92.11, IS USED AT DEPTH 16. APPEARS TO GIVE AT LEAST 13 SIGNI-
[4]   ⍝FICANT FIGURES. PORTED TO MAINFRAME 1719; LINE [8] HAS BEEN CHANGED TO
[5]   ⍝ AVOID UNDERFLOW PROBLEMS WITH ×\
[6]    →((⍴,A←(|X<4)/Z←X←,X)=⍴,X)/3+⎕LC
[7]    B←(|X≥4)/X
[8]    →((⍴,B)=⍴X)/BIG
[9]    A←0.5+(÷A×((*A*2)×○2)*0.5)×+/×\(A*2)∘.÷¯1+2×⍳(C←10⌈⌈10×⌈/|A)+1
[10]   Z[(|X<4)/⍳⍴X]←A
[11]   →((⍴,A)=⍴X)/0
[12]  BIG:C← 16589790 56295540 52050600 19934640 3680160 341952 15232 256
[13]   D← 2027025 32432400 75675600 60540480 21621600 3843840 349440 15360 256
[14]   B←1-(B÷2×((○2)×*B*2)*0.5)×(+/((0.5×,B*2)∘.*0,⍳7)×((⍴,B),8)⍴C)÷+/((0.5×,B*
       2)∘.*0,⍳8)×((⍴,B),9)⍴D
[15]   Z[(|X≥4)/⍳⍴X]←B
      ∇
```

```
      ∇ Z←NORMPTH P;A;B;C;D;Q;T;S;R;F
[1]   ⍝ IMPLEMENTS ALGORITHM AS 111 BY BEASLEY   SPRINGER, APPLIED STAT, 1977
[2]   ⍝ FOR A VECTOR INPUT OF FRACTIONS, RETURNS CORRESPONDING NORMAL QUAN-
[3]   ⍝TILES CLAIMED ACCURACY IS BETTER THAN 1.5×10* 8. FOR GREATER ACCURACY,
[4]   ⍝ESPECIALLY FOR EXTREME P VALUES, ADD ONE OR MORE NEWTON-RAPHSON LOOPS.
[5]    →(∨/(P≤0),(P≥1))/ERR
[6]    →(∨/((|Q←,P-0.5)≤0.42))/3+⎕LC
[7]    S←Z←,Q
[8]    →EXT
[9]    T←(0.42≥|Q)/Z←,Q
[10]   →(F←((⍴,T)=⍴,P))/2+⎕LC
[11]   S←(0.42<|Q)/Q
[12]   A← 2.50662823884 ¯18.6150006252 41.39119773534 ¯25.44106049637
[13]   B←  8.4735109309 23.08336743743 ¯21.06224101826 3.13082909833
[14]   T←T×(((T*2)∘.*0,⍳3)+.×A)÷1+((T*2)∘.*⍳4)+.×B
[15]   Z[(0.42≥|Q)/⍳⍴,Q]←T
[16]   →(F=1)/0
[17]  EXT:C←  2.78718931138 ¯2.29796479134 4.85014127135 ¯2.32121276858
[18]   D← 3.54388924762 1.63706781897
[19]   S←(×S)×((R∘.*0,⍳3)+.×C)÷1+(((R←(|⍟0.5-|S)*0.5)∘.* 1 2)+.×D)
```

```
[20]  Z[(0.42<|Q)/⍳⍴Q]←S
[21]  →0
[22] ERR:'ONE OR MORE P VALUES ARE OUT OF RANGE.'
      ∇


      ∇ P←PERM N;X;Y;Z
[1]   ⍝ THIS FUNCTION IS CALLED BY SPEARP. IT GENERATES ALL POSSIBLE
[2]   ⍝ PERMUTATIONS OF N RANKS. N=SAMPLE SIZE.
[3]    →0×⍳N=P← 1 1 ⍴1
[4]    Z←PERM N-1
[5]    P←⍳X←0
[6]   L1:→0×⍳N<X←X+1
[7]    Y←(~(⍳N)∊X)\Z
[8]    Y[;X]←N
[9]    P←((X×!N-1),N)⍴(,P),,Y
[10]   →L1
      ∇


      ∇ M←N QUANC Q;I;J;K;L;M;U
[1]   ⍝  THIS FUNCTION GIVES A CHOICE OF NONPARAMETRIC CONFIDENCE INTERVALS
[2]   ⍝ FOR THE QTH QUANTILE OF A CONTINUOUS POPULATION.  N=SAMPLE SIZE,
[3]   ⍝ Q=QUANTILE. IT CALLS THE SUBPROGRAM CBIN.
[4]
[5]   ⍝ CENTERS CHOICES AT THE ORDER STATISTIC NEAREST ESTIMATE.
[6]    I←⌊0.5+N×Q÷100
[7]   ⍝ J IS 1.5 STANDERD DEVIATIONS(APPROX.)
[8]    J←⌊0.5+1.5×(I×1-Q÷100)*0.5
[9]    I←I+J× ¯1 1
[10]  ⍝ RUN OUT K BOTH WAYS.
[11]   K←¯1+⍳1+⌊/(I[2],N-I[1])
[12]  ⍝ CONFIDENCE COEFFICIENTS
[13]   L←-/⍉(2,⍴K)⍴(N,0.01×Q) CBIN((I[1]+K-1),(I[2]-K))
[14]   M←'ORDER STATISTICS|COEFFICIENTS'
[15]  ⍝ PART OF FORMATTING OUTPUT
[16]   M←M,[0.1] 29⍴((16⍴'-'),'|')
[17]   U← 7 0 5 0 ⍕⍉(2,⍴K)⍴(1+((I[2]-K),I[1]+K-1))
[18]   U←U,((⍴K),4)⍴'   |'
[19]   M←M,[1] U, 12 6 ⍕((⍴K),1)⍴L
      ∇


      ∇ SPN←N SPAPROX X;Y
[1]   ⍝   THIS FUNCTION IS A SUBPROGRAM OF SPEARMAN'S R (SPMANP)
[2]   ⍝ IT APPROXIMATES THE CUMMULATIVE PROB FOR R WHEN PASSED THE SAMPLE
[3]   ⍝ SIZE IN THE LEFT ARGUMENT AND THE ABSOLUTE VALUE OF R IN THE RIGHT
[4]   ⍝ ARGUMENT.  SUBPROGRAMS OF THIS FUNCTION INCLUDE: TDISTN
[5]
[6]   ⍝ CALCULATE THE CONTINUITY CORRECTION
[7]    Y←6÷N×¯1+N*2
[8]   ⍝ TRANSFORM THE STATISTIC R INTO ONE THAT CAN BE USED WITH THE STUDENT
[9]   ⍝ T DISTRIBUTION
[10]   X←(X-Y)×((N-2)÷1-(X-Y)*2)*0.5
[11]  ⍝ CALL THE T DIST FUNCTION TO CALCULATE THE P-VALUE
[12]   SPN←1-(N-2) TDISTN X
      ∇


      ∇ SPEAR←SPEARP N;C1;A;B;C;D;E;M;N;D1;D2;LIM;R;CDF
[1]   ⍝   THIS FUNCTION IS A SUBPROGRAM OF SPEARMAN'S R (SPMANP). IT
[2]   ⍝ CALCULATES THE EXACT CUMULATIVE DIST. FOR R FOR THE SAMPLE SIZE
[3]   ⍝ PASSED AS THE RIGHT ARGUMENT. BECAUSE OF THE LARGE COMPUTER MEMORY
[4]   ⍝ REQUIREMENTS, N IS LIMITED TO SIX ON THE PC AND 7 ON THE MAINFRAME.
[5]   ⍝ SUBPROGRAMS CALLED BY THIS FUNCTION INCLUDE: PERM
[6]    A←B←0
[7]   ⍝ INITIALIZE VARIABLES, VECTORS, AND MATRICES.
[8]    C1←D1←⍳0
[9]    M←(N,N)⍴0
[10]   C←0,⍳(N-1)
[11]  ⍝ THIS LOOP GENERATES AN N×N ARRAY OF THE POSSIBLE VALUES OF DIFFERENCES
[12]  ⍝ BETWEEN ANY TWO PAIRED RANKS BETWEEN SAMPLES.
[13] L1:B←B+1
[14]   M[B;]←C-A
[15]   A←A+1
[16]   →(B<N)/L1
[17]  ⍝ NOW CALCULATE THE SQUARES OF ALL POSSIBLE DIFFERENCES.
[18]   M←M*2
[19]  ⍝  CALL PERM TO LIST ALL POSSIBLE PERMUTATIONS OF N NUMBERS
[20]   D←PERM N
[21]  ⍝ CALCULATE SIZE LIMIT OF FINAL VECTOR OF R STATS.,
[22]   LIM←1+(((N*3)-N)÷12)
[23]  ⍝ INITIALIZE VALUES BEFORE INDEXING OUT COMBINATIONS OF ALL POSSIBLE
```

```
[24]  ⍝ SQUARED VALUES.
[25]   A←0
[26]   E←(N,!N)ρ0
[27]  ⍝ THIS LOOP CALCULATES ALL POSSIBLE COMBINATIONS OF THE SQUARED VALUES.
[28]  L2:A←A+1
[29]   E[A;]←M[A;D[;A]]
[30]   →(A<N)/L2
[31]  ⍝ ADD DOWN ALL ROWS FOR EACH COLUMN TO SUM UP SQUARES COMBINATIONS.
[32]   D2←+⌿E
[33]  ⍝ ADD UP NUMBER OF DUPLICATED SUMS OF D-SQUARED VALUES AND COMPRESS
[34]  ⍝ VECTOR DOWN TO UNIQUE VALUES.
[35]  L3:C1←C1,(⌊/D2)
[36]   D1←D1,(+/((⌊/D2)=D2))
[37]   D2←((⌊/D2)≠D2)/D2
[38]   →((ρC1)<LIM)/L3
[39]  ⍝ TRANSFORM SUM OF SQUARES VALUES TO SPEARMAN'S R STATISTIC.
[40]   R←1-(6×C1)÷(N×((N*2)-1))
[41]  ⍝ CALCULATE CDF VALUES ASSOCIATED WITH THE R STATISTIC.
[42]   CDF←(+\D1)÷!N
[43]  ⍝ FORM TWO ROW MATRIX FOR OUTPUT OF R STATS AND CDF VALUES.
[44]   SPEAR←(2,(ρC1))ρ(R,CDF)
       ∇



      ∇ SPM←X SPMANP Y;C;D;DD;D1;D2;N;DENOMR;XX;Y1;NS;NUMR;P;PV;PVAL;SU;SV;RHO;
        U;V;ARHO;U1;V1;X;Y;WW
[1]   ⍝       THIS FUNCTION IS A SUBPROGRAM OF NONPAR LINEAR REGRESSION (NPLR)
[2]   ⍝ AND SPEARMAN'S R (SPMAN).  IT COMPUTES THE SPEARMAN R STATISTIC
[3]   ⍝ AND ASSOCIATED P-VALUES. THE LEFT ARGUMENT THAT IS PASSED IS THE X
[4]   ⍝ SAMPLE; THE RIGHT ARGUMENT IS THE Y SAMPLE.
[5]   ⍝ SUBPROGRAMS OF THIS FUNCTION INCLUDE: TIES, TIESK, SPEARP, SPAPROX,
[6]   ⍝ INTERP, AND THE VARIABLE PMATSP.
[7]
[8]   ⍝            ORDER Y IN INCREASING ORDER OF X
[9]    V←Y[⍋X]
[10]  ⍝           ORDER X IN INCREASING ORDER
[11]   U←X[⍋X]
[12]  ⍝             COMPUTE CURRENT RANKING OF Y
[13]   C←⍋⍋V
[14]  ⍝             NOW ORDER Y RANKS IN INCREASING ORDER
[15]   D1←V[⍋V]
[16]  ⍝      IF TIES EXIST IN EITHER X OR Y RANKED VECTOR USE MID-RANK METHOD
[17]   DD←1 TIES D1
[18]   XX←1 TIES U
[19]  ⍝         FIND ORIGINAL RANKING OF Y WITH TIES RESOLVED
[20]   Y1←DD[C]
[21]  ⍝ RECORD SIZE OF INPUT VECTOR
[22]   N←ρX
[23]  ⍝ CALCULATE DIFFERENCES BETWEEN RANKS OF X AND Y VECTORS
[24]   D←XX-Y1
[25]  ⍝ DETERMINE THE SUN OF SQUARES OF THE DIFFERENCES
[26]   D2←+/(D*2)
[27]  ⍝     OBTAIN THE NUMBER OF TIES IN EACH VECTOR USING THE TIESK FUNCTION
[28]   U1←TIESK U
[29]   V1←TIESK D1
[30]   SU←((+/(U1*3))-(+/U1))÷12
[31]   SV←((+/(V1*3))-(+/V1))÷12
[32]   NS←N×((N*2)-1)
[33]  ⍝ CALCULATE THE R STATISTIC INCLUDING THE CORRECTION FOR TIES
[34]   NUMR←(NS)+((¯6)×D2)+((¯6)×(SU+SV))
[35]   DENOMR←((NS-(12×SU))*0.5)×((NS-(12×SV))*0.5)
[36]   RHO←NUMR÷DENOMR
[37]   ARHO←|RHO
[38]   →(N≥6)/L1
[39]  ⍝         CALL SPEARP TO CALCULATE THE RIGHT TAIL OF THE CDF OF R
[40]   P←SPEARP N
[41]   →L2
[42]  L1:→(N>10)/L3
[43]   P←PMATSP[(N-5);;]
[44]  ⍝  CHANGE SIZE OF P TO AN M×N MATRIX
[45]   P←P[1;;]
[46]  ⍝    CALL INTERP TO CALCULATE P-VALUE BY INTERPOLATION
[47]  L2:PVAL←ARHO INTERP P
[48]   →(PVAL=¯1)/L7
[49]   PVAL←0.5
[50]   →L7
[51]  ⍝ CALCULATE P VALUE USING STUDENT T APPROX.
[52]  L3:PVAL←N SPAPROX ARHO
[53]  L7:SPM←(RHO),PVAL
       ∇
      ∇ P←K TDISTN X;V
[1]   ⍝ THIS FUNCTION IS A SUBPROGRAM OF THE CUMULATIVE PROBABILITY GENERATOR
[2]   ⍝ FOR SPEARMAN'S R (SPEARP).  IT CALCULATES THE CDF AT X USING
[3]   ⍝ THE STUDENT'S T DIST WITH K DEGREES OF FREEDOM.
[4]   ⍝ THIS FUNCTION CALLS ON THE  'F' DISTRIBUTION FUNCTION (FDISTN).
[5]    V←(X←,X)≥0
[6]    P←0.5×(1,K) FDISTN X*2
[7]    →8×ι0=∨/V
[8]    P[V/ιρX]←0.5+V/P
[9]    →0×ι1=∧/V
[10]   P[(~V)/ιρX]←0.5-(~V)/P
      ∇


      ∇ TI←BB TIES B;C;D;I;N;T;Y;Z;K;M;L;PP;NR;MM
[1]   ⍝     THIS FUNCTION IS A SUBPROGRAM OF KENDALL'S B (KEN), SPEARMAN'S
[2]   ⍝ R (SPMANP), KRUSKAL-WALLIS (KRWL), MANN-WHITNEY (MANW)
[3]   ⍝ AND WILCOXON (WISIG).  IT CHECKS THE RIGHT ARG. VECTOR FOR TIES AND
[4]   ⍝ CHANGES THE TIED POSITIONS OF THE LEFT ARG. BY THE MIDRANK METHOD.
[5]
[6]    N←ρB
[7]   ⍝          IF NO VECTOR OF RANKS IS PASSED; GENERATE ONE
[8]    →((ρρBB)≠0)/L6
[9]    BB←ιN
[10]  L6:I←ρ0
[11]   L←Nρ0
[12]   T←1
[13]  ⍝            CHECKING FOR TIES BY INCREMENTING THRU THE VECTOR
[14]  L3:C←(T↓B)=B[T]
[15]  ⍝            COUNT NUMBER OF TIES; IF NO TIES GO TO L2
[16]   D←+/C
[17]   →(D=0)/L2
[18]  ⍝  RECORD WHERE TIES STARTED AND HOW MANY RANKS INVOLVED
[19]   I←I,T,(D+1)
[20]  ⍝         INCREMENT NEXT T BY THE NUMBER OF TIES ENCOUNTERED PLUS 1
[21]  L2:T←T+(D+1)
[22]  ⍝IF T LESS THAN SIZE OF ORIGINAL VECTOR GO TO L3 AND START AGAIN AT NEW T
[23]   →(T<N)/L3
[24]   Y←ρI
[25]  ⍝             ASSIGN THE RANKS OF THE LEFT ARG. TO TI
[26]   TI←BB
[27]  ⍝                    IF NO TIES FOUND QUIT
[28]   →(Y=0)/0
[29]   Z←0
[30]  ⍝ LOCATE THE INDEXED POSITIONS OF TIED RANKS
[31]  L5:PP←((I[1+Z])-1)↓ι((I[1+Z]+I[2+Z])-1)
[32]  ⍝ FIND THE MIDRANK VALUE OF THESE RANKS
[33]   NR←(+/TI[PP])÷I[2+Z]
[34]  ⍝  SET UP VECTOR WITH ZEROS AND ONES; ONES WHERE TIE RANKS INVOLVED
[35]   K←0
[36]   MM←Nρ0
[37]  L4:K←K+1
[38]   M←(TI=TI[PP[K]])
[39]   MM←MM+M
[40]   →(K<I[2+Z])/L4
[41]  ⍝ SET UP A VECTOR WITH ZEROS WHERE TIED RANKS OCCUR
[42]   L←~MM
[43]  ⍝ TRANSFORM ONES OF MM VECTOR TO MIDRANK VALUE
[44]   MM←NR×MM
[45]  ⍝ TRANSFORM ONES OF L VECTOR TO REMAINING UNCHANGED RANKING VECTOR
[46]   L←TI×L
[47]  ⍝ FILL IN MIDRANK VALUES
[48]   TI←L+MM
[49]   Z←Z+2
[50]  ⍝   DO THE SAME FOR ANY OTHER TIES INVOLVED BUT WITH NEWLY COMPUTED TI
[51]   →(Y>Z)/L5
      ∇


      ∇ TIE←TIESK AA;AA;B;C;D;I;N;T
[1]   ⍝     THIS FUNCTION IS A SUBPROGRAM OF KENDALL'S B (KEN), SPEARMAN'S
[2]   ⍝ R (SPMAN) AND (SPMAN1), AND KRUSKAL-WALLIS (KRWL).  IT CHECKS THE
[3]   ⍝ RIGHT ARGUMENT FOR TIES AND RECORDS THE NUMBER OF OCCURENCES OF EACH
[4]   ⍝ TIE AND THE TOTAL NUMBER OF TIES IN THE VECTOR.
[5]
[6]   ⍝       ASSIGN ORDERED  VECTOR  TO B AND INITIALIZE VALUES
[7]    N←ρAA
[8]    B←AA[⍋AA]
[9]    TIE←ρ0
[10]   T←1
```

```
 ⍝         CHECKING FOR TIES BY INCREMENTING THRU THE VECTOR
L3:C←(T↓B)=B[T]
 ⍝         COUNT NUMBER OF TIES; IF NO TIES GO TO L2
 D←+/C
 →(D=0)/L2
 TIE←TIE,(D+1)
 ⍝      INCREMENT NEXT T BY THE NUMBER OF TIES ENCOUNTERED PLUS 1
L2:T←T+(D+1)
 ⍝IF T LESS THAN SIZE OF ORIGINAL VECTOR GO TO L3 AND START AGAIN AT NEW T
 →(T<(N-1))/L3
 ∇


 ∇ VAR←VARMW B;C;D;E;D1;E1
 ⍝      THIS FUNCTION IS A SUBPROGRAM OF THE MANN-WHITNEY TEST (MANW).
 ⍝ IT GENERATES THE RANKING SCHEME USED IN CALCULATING THE
 ⍝ DIFFERENCES IN SCALE (1 ASSIGNED SMALLEST, 2 ASSIGNED LARGEST, 3
 ⍝ ASSIGNED NEXT LARGEST, 4 SECOND SMALLEST,ETC. BY TWOS TILL PROPER
 ⍝ SAMPLE SIZE IS REACHED). SAMPLE SIZE IS PASSED IN THE RIGHT ARG.
 E←E1←⍴0
 D←0
 ⍝         FIND FLOOR OF MIDPOINT OF VECTOR AND ASSIGN TO C
 C←⌊(B÷2)
 ⍝ LOOPS GENERATE RANKING VALUES LEFT HALF FIRST
L2:D←D+1
 E←E,D
 →((⍴E)=C)/L3
 D←D+3
 E←E,D
 →((⍴E)<C)/L2
 ⍝              NOW GENERATE RIGHT HALF
L3:D1←1
 D1←D1+1
 E1←E1,D1
 →((⍴E1)=C)/L5
L6:D1←D1+1
 E1←E1,D1
 →((⍴E1)=C)/L5
 D1←D1+3
 E1←E1,D1
 →((⍴E1)<C)/L6
 ⍝ IF SIZE OF VECTOR IS ODD VALUE CONCATENATE MIDDLE RANK IN BETWEEN HALFS
L5:→((2|B)≠0)/L7
 ⍝ IF SIZE IS EVEN CONCATENATE LEFT HALF WITH THE REVERSE OF THE RIGHT
 VAR←E,(⊖E1)
 →0
L7:VAR←E,B,(⊖E1)
 ∇


 ∇ WIL←WILP NN;N;A;P;T;NN;W;PP;NM
 ⍝    THIS FUNCTION IS A SUBPROGRAM OF THE WILCOXON SIGNED RANK TEST
 ⍝ (WISIG). IT GENERATES THE CUMULATIVE DIST. FOR THE TEST STATISTIC
 ⍝ THE GENERATOR USES A RECURSIVE FORMULA. NN=SAMPLE SIZE.
 NM←(⌊((+/⍳NN)÷2))+1
 N←2
 ⍝        SET P EQUAL TO PROB. DIST. WHEN N EQUALS 2.
 P←4⍴1
L3:N←N+1
 A←1
 ⍝         SET T VECTOR TO PROPER SIZE OF ZEROS.
 T←(+/(⍳N))⍴0
 ⍝         IF A≤N USE TRUNCATED RELATION TO COMPUTE OCCURRENCES.
L4:→(A≤N)/L1
 ⍝        IF A>N USE FULL FORMULA TO COMPUTE OCCURRENCES.
 →(A>N)/L2
 ⍝ WHILE (A-N) IS NEGATIVE TRUNCATE FORMULA  TO AVOID A NEGATIVE INDEX.
L1:T[A]←P[A]
 A←A+1
 →L4
 ⍝      IF A IS LARGER THAN THE LENGTH OF P GO TO L6.
L2:→(A>(⍴P))/L6
 ⍝     ONCE (A-N) BECOMES POSITIVE: THE RECURSIVE FORMULA CAN BE USED.
 T[A]←P[(A-N)]+P[A]
 →L7
 ⍝     ONCE A IS LARGER THAN THE LENGTH OF P TRUNCATE FUNCTION AGAIN.
L6:T[A]←P[(A-N)]
L7:A←A+1
 ⍝     IF A AS AN INDEX HAS NOT EXCEEDED N(N+1)/2 GO AGAIN.
 →(A≤(+/(⍳N)))/L4
 ⍝ CONVERT T INTO P AND CONCATENATE 1 FOR USE IN NEXT ITERATION OR OUTPUT.
L5:P←((+/(⍳N))⍴T),1
 PP←NM↑P
 WIL←(+\PP)÷(2*NN)
 ⍝    CHECK IF LENGTH OF INPUT VECTOR EXCEEDS NUMBER OF N'S GENERATED.
 →(NN>N)/L3
 ∇
```

APPENDIX G

LISTINGS OF PROGRAMS USED TO GENERATE C.D.F. COMPARISON TABLES

```
    ∇ KENTEST;N;ALPHA;B;A;C;TAU;P;NUM;DEN;Z;ZZ;ERRZZ;M;ZZC;D;AA;PP;NUMC;I;H;F
      ;FS;S;J;ERRZZC;ZC;KK

 ⍝THIS PROGRAM GENERATES TABLES OF C.D.F. COMPARISONS FOR KENDALL'S B
 ⍝          SET SAMPLE SIZE AND ALPHA VALUES.
  N←11
  ALPHA← 0.01 0.02 0.03 0.05 0.08 0.13 0.18
 ⍝          THIS LOOP INCREMENTS SAMPLE SIZE.
 L1:PP←⍴0
  M←⍴0
  KK←⍴0
  FS←⍴0
  ZZ←⍴0
  ZZC←⍴0
  N←N+1
  D←0
  B←0
 ⍝         COMPUTE CUMULATIVE DIST AND ASSOCIATED STATS.
  P←KENDALP N
 ⍝         THIS LOOP CALCULATES ALPHA VALUES AND APPORXIMATIONS.
 L2:B←B+1
  C←ALPHA[B]
  A←(+/(P[2;]≤C))
  TAU←P[1;A]
  KK←KK,TAU
  AA←6÷N×¯1+N*2
  PP←PP,P[2;A]
 ⍝         COMPUTE NORMAL APPROXIMATION.
  NUM←3×(TAU)×((N×(N-1))*0.5)
  DEN←(2×((2×N)+5))*0.5
  Z←NUM÷DEN
  ZZ←ZZ,(1-NORMCDF Z)
 ⍝         COMPUTE NORMAL APPROXIMATION WITH CONTINUITY CORRECTION.
  NUMC←3×(TAU-AA)×((N×(N-1))*0.5)
  ZC←NUMC÷DEN
  ZZC←ZZC,(1-NORMCDF ZC)
  →(B<7)/L2
 ⍝          COMPUTE ERROR DIFFERENCES.
  ERRZZ←PP-ZZ
  ERRZZC←PP-ZZC
 ⍝          PRINT OUT TABLE OF VALUES.
  I←0.1
  J←195
  S←'TEST STAT. VALUE  '
  M←M,'                    'PROB[B ≥ B̲]; FOR SAMPLE SIZE EQUAL TO ',(2 0 ⍕N),'
  .
  →L3
 L5:J←194
 L3:M←M,[I] ⎕AV[(18⍴197),J,61⍴(8⍴197),J]
  →(D=4)/L4
  H← 1 18 ⍴S
  →(D≠0)/L6
  F←'K4G<| 9.9999 >' ⎕FMT(1 7 ⍴KK)
  F← 1 62 ⍴F
  FS←H,F
  →L7
 L6:FS←⍴0
  H←17⍴S
  FS←FS,H
  C←0
 L9:C←C+1
  F←' |',(7 5 ⍕PP[C])
  FS←FS,F
  →(C<7)/L9
 L7:M←M,[1] FS
  D←D+1
  I←1
  →(N1,N2,N3,L5)[D]
 N1:S←'EXACT C.D.F.       '
  J←198
  →L3
 N2:S←'ERROR; NORMAL      '
  PP←ERRZZ
  →L3
```

```
N3:S←'ERROR: NORM. W/CC '
 PP←ERRZZC
 →L3
L4:M
 →(N<14)/L1
 ∇


  ∇ KWTEST;A;B;C;PP;PC;NN;ALPHA;N;K;P;F;H;PVALUE;PVAL;PF1;PF;PVF;PVF1;D;ERR
    H;ERRF;ERRF1;CDF;CC;KK;I;J;FF;S;M

 ⍝ THIS PROGRAM GENERATES TABLES OF C.D.F. COMPARISONS FOR THE KRUSKAL-
 ⍝ WALLIS TEST.

  A← 4 4 4
  CC←2
  ALPHA← 0.01 0.02 0.03 0.05 0.08 0.13 0.18
  →L2
 L1:A←NN
 ⍝ CALL KRUWALP TO GENERATE EXACT DISTRIBUTION FOR SELECTED SAMPLE SIZE.
 L2:P←KRUWALP A
  B←0
  N←+/A
  CC←CC+1
  K←⍴A
  M←⍴0
  PP←⍴0
  KK←⍴0
  PC←⍴0
  PF←⍴0
  PF1←⍴0
  CDF←P[2;]
 L4:B←B+1
 ⍝ LOCATE POSITION OF EXACT CDF VALUE ≤ ALPHA.
  D←(+/(CDF≤ALPHA[B]))
  →(D=0)/L3
  D←1
 ⍝ DETERMINE CORRESPONDING TEST STATISTIC VALUE.
 L3:H←P[1;D]
  KK←KK,H
 ⍝ RECORD EXACT VALUE OF CDF.
  PVALUE←P[2;D]
  PP←PP,PVALUE
 ⍝ COMPUTE CORRESPONDING P-VALUE USING CHISQ APPROX.
  PVAL←(K-1) CHISQ H
  PC←PC,(1-PVAL)
 ⍝COMPUTE CORRESPONDING P-VALUE USING  F APPROXIMATION
  F←((N-K)×H)÷((K-1)×((N-1)-H))
  PVF←((K-1),(N-K)) FDISTN F
  PF←PF,(1-PVF)
 ⍝ COMPUTE CORRESPONDING P-VALUE USING F APPROX W/ 1 LESS D.F. IN DENOM
  PVF1←((K-1),((N-K)-1)) FDISTN F
  PF1←PF1,(1-PVF1)
  →(B<7)/L4
  D←0
  ERRH←PP-PC
  ERRF←PP-PF
  ERRF1←PP-PF1
 ⍝ PRINT OUT TABLE OF VALUES.
  I←0.1
  J←'-'
  S←'TEST STAT. VALUE '
  M←M,'      PROB[H ≥ ',⎕AV[46],']; FOR A GROUP OF 3 SAMPLES CONSISTING OF 4
  , 4, AND ',(⍕CC),' OBS.          '
  →L8
 L5:J←'-'
 L8:M←M,[I](17⍴'-'),J,61⍴(8⍴'-'),J
  →(D=5)/L9
  F←⍴0
  H←16⍴S
  F←F,H
  C←0
  →(D≠0)/L10
 L11:C←C+1
  FF←' |',(7 4 ⍕KK[C])
  F←F,FF
  →(C<7)/L11
  →L12
 L10:C←C+1
  FF←' |',(7 5 ⍕KK[C])
  F←F,FF
  →(C<7)/L10
 L12:M←M,[1] F
  D←D+1
  I←1
  →(N1,N2,N3,N4,L5)[D]
 N1:S←'ALPHA VALUE       '
  KK←PP
  J←'+'
 →L8
N2:S←'ERROR; CHISQUARE'
 KK←ERRH
 →L8
N3:S←'ERROR; F DIST    '
 KK←ERRF
 →L8
N4:S←'ERROR; F W/¯1 DF'
 KK←ERRF1
 →L8
L9:M
 →(CC<3)/L1
 ∇


 ∇ KWTESTSM;A;R2;PC;NN;ALPHA;N;K;Y;W;H;YY;PVAL;PF1;HH;MM;M;X;ERRC;ERRF1;F;
   PV;B;KK;D;PVF1;Q;FF;U;C;I;J;S

⍝ THIS PROGRAM GENERATES C.D.F. COMPARISONS FOR THE KRUSKAL-WALLIS TEST
⍝ BASED ON 30000 RANDOM VECTORS OF 40 RANKINGS WHICH SIMULATE COMPARING
⍝ 5 SAMPLES OF 8 OBSERVATIONS EACH.

 A←1
 W←10000
 K←5
 HH←⍴0
 N←40
 NN←0
L1:NN←NN+1
 Q←8
 MM←0
⍝ GENERATE A RANDOM VECTOR OF RANKINGS.
 X←40?40
 YY←⍴0
⍝ SUM THE RANKS OF EACH SAMPLE.
L2:Y←+/(MM↓(Q↑X))
 YY←YY,Y
 Q←Q+8
 MM←MM+8
 →(Q≤40)/L2
 R2←+/((YY*2)÷8)
⍝ CALCULATE THE H STATISTIC FOR EACH VECTOR OF RANKINGS.
 H←((3÷410)×R2)-123
 HH←HH,H
 →((NN÷10000)=1,2,3,4)/L4
⍝ CHANGE RANDOM SEED EVERY 10000 ITERATIONS.
 ⎕RL← 0 60 60 1000 ⊥ ¯4↑⎕TS
L4:→(NN<W)/L1
⍝ ORDER ALL H'S FROM LARGEST TO SMALLEST.
 HH←HH[⍒HH]
 B←0
 D←0
 M←⍴0
 ALPHA← 0.01 0.02 0.03 0.05 0.08 0.13 0.18
 KK←⍴0
 PF1←⍴0
 PC←⍴0
L3:B←B+1
⍝ DETERMINE DESIRED H VALUE CORRESPONDING TO ALPHA VALUE.
 H←HH[(NN×ALPHA[B])]
 KK←KK,H
⍝ COMPUTE CORRESPONDING P-VALUE USING CHISQ APPROX.
 PVAL←(K-1) CHISQ H
 PC←PC,(1-PVAL)
⍝ COMPUTE CORRESPONDING P-VALUE USING F APPROX W/ 1 LESS D.F. IN DENOM
 F←((N-K)×H)÷(4×(39-H))
 PVF1←((K-1),((N-K)-1)) FDISTN F
 PF1←PF1,(1-PVF1)
 →(B<7)/L3
 ERRC←ALPHA-PC
 ERRF1←ALPHA-PF1
⍝           PRINT OUT TABLE OF VALUES.
 I←0.1
 J←'-'
 S←'TEST STAT. VALUE '
 M←M,'    PROB[H ≥ ',⎕AV[46],'];  BASED ON ',(5 0 ⍕W),' GENERATED H''S FOR 5
  SAMPLES OF 8 OBS. EACH.
 →L8
L5:J←'-'
L8:M←M,[I](17⍴'-'),J,61⍴(8⍴'-'),J
 →(D=4)/L9
 F←⍴0
 H←16⍴S
 F←F,H
 C←0
 →(D=0)/L10
L11:C←C+1
 FF←' |',(7 3 ⍕KK[C])
 F←F,FF
```

```
  →(C<7)/L11
  →L12
 L10:C←C+1
  FF←' | ',(7 5 ⍕KK[C])
  F←F,FF
  →(C<7)/L10
 L12:M←M,[1] F
  D←D+1
  I←1
  →(N1,N2,N3,L5)[D]
 N1:S←'ALPHA VALUE       '
  KK←ALPHA
  J←'+'
  →L8
 N2:S←'ERROR; CHISQUARE'
  KK←ERRC
  →L8
 N3:S←'ERROR; F W/¯1 DF'
  KK←ERRF1
  →L8
 L9:M
  U← 20000 30000 1
  W←U[A]
  A←A÷1
  →(A<4)/L1
  ∇


  ∇ MANTEST;U;NN;MM;ALPHA;B;A;C;T;P;NUM;DEN;Z;ZZ;ERRZZ;M;ZZC;D;PP;NUMC;I;E;
    F;FS;S;J;TT;T;TTC;TC;NUMC;NUMT;DENOM;DENOMC;ERRAVE;ERRTT;ERRTC;ERRZZC;Z
    C;ERRTCZC;KK

 ⍝ THIS PROGRAM GENERATES TABLES OF C.D.F. COMPARISONS FOR THE MANN-
 ⍝ WHITNEY TEST.

 ⍝            SET SAMPLE SIZE AND ALPHA VALUES.
  NN←7
  MM←7
  ALPHA← 0.01 0.02 0.03 0.05 0.08 0.13 0.18
 ⍝            THIS LOOP INCREMENTS SAMPLE SIZE.
 L1:PP←⍴0
  M←⍴0
  KK←⍴0
  FS←⍴0
  ZZ←⍴0
  ZZC←⍴0
  TT←⍴0
  TTC←⍴0
  NN←NN+1
  MM←MM+1
  D←0
  B←0
 ⍝          COMPUTE CUMULATIVE DIST AND ASSOCIATED STATS.
  P←NN MANWP MM
 ⍝          THIS LOOP CALCULATES ALPHA VALUES AND APPORXIMATIONS.
 L2:B←B+1
  C←ALPHA[B]
  A←(+/(P≤C))
  U←A-1
  KK←KK,U
  PP←PP,P[A]
 ⍝          COMPUTE NORMAL APPROXIMATION.
  NUM←U-((NN×MM)÷2)
  DEN←((NN×MM×(NN+MM+1))÷12)*0.5
  Z←NUM÷DEN
  ZZ←ZZ,NORMCDF Z
 ⍝          COMPUTE NORMAL APPROXIMATION WITH CONTINUITY CORRECTION.
  NUMC←(U+0.5)-((NN×MM)÷2)
  ZC←NUMC÷DEN
  ZZC←ZZC,NORMCDF ZC
 ⍝         COMPUTE STUDENT T APPROXIMATION
  DENOM←(((NN+MM-1)-(Z*2))÷(NN+MM-2))*0.5
  T←Z÷DENOM
  TT←TT,(NN+MM-2) TDISTN T
 ⍝     COMPUTE STUDENT T APPROXIMATION WITH CONTINUITY CORRECTION.
  NUMT←(NUM)+0.5
  DENOMC←((((NN+MM-1)×(DEN*2))÷(NN+MM-2))-(((NUM+0.5)*2)÷(NN+MM-2)))*0.5
  TC←NUMT÷DENOMC
  TTC←TTC,((NN+MM-2) TDISTN TC)
  →(B<7)/L2
 ⍝           COMPUTE ERROR DIFFERENCES.
  ERRZZ←PP-ZZ
  ERRZZC←PP-ZZC
  ERRTT←PP-TT
  ERRTC←PP-TTC
  ERRAVE←PP-((ZZ+TT)÷2)
  ERRTCZC←PP-((ZZC+TTC)÷2)
 ⍝           PRINT OUT TABLE OF VALUES.
  I←0.1
  J←195
 S←'TEST STAT. VALUE   '
 M←M,'          PROB[U ≤ U]; FOR SAMPLE SIZES N EQUAL TO ',(2 0 ⍕NN),' AND
  M EQUAL TO ',(2 0 ⍕MM),'.                   '
  →L3
L5:J←194
L3:M←M,[I] ⎕AV[(18⍴197),J,61⍴(8⍴197),J]
 →(D=8)/L4
 H← 1 18 ⍴S
 →(D=0)/L6
 F←'BG<|  ZZ9   >' ⎕FMT(1 7 ⍴KK)
 F← 1 62 ⍴F
 FS←H,F
 →L7
L6:FS←⍴0
 H←17⍴S
 FS←FS,H
 C←0
L9:C←C+1
 F←' |',(7 5 ⍕PP[C])
 FS←FS,F
 →(C<7)/L9
L7:M←M,[1] FS
 D←D+1
 I←1
 →(N1,N2,N3,N4,N5,N6,N7,L5)[D]
N1:S←'EXACT C.D.F.       '
 J←198
 →L3
N2:S←'ERROR; NORMAL      '
 PP←ERRZZ
 →L3
N3:S←'ERROR; NORM. W/CC'
 PP←ERRZZC
 →L3
N4:S←'ERROR; T DIST      '
 PP←ERRTT
 →L3
N5:S←'ERROR; T W/CC      '
 PP←ERRTC
 →L3
N6:S←'ERROR; AVE T/Z     '
 PP←ERRAVE
 →L3
N7:S←'ERROR; AVE TC/ZC '
 PP←ERRTCZC
 →L3
L4:M
 →(NN<9)/L1
∇
```

```
∇ SIGNTEST;N;ALPHA;B;A;C;K;P;Z;ZZ;ERRZZ;M;ZZC;D;PP;I;H;F;FS;S;J;ERRZZC;ZC
 ;KK

⍝ THIS PROGRAM GENERATES TABLES OF C.D.F. COMPARISONS FOR THE SIGN TEST.

⍝          SET SAMPLE SIZE AND ALPHA VALUES.
 N←23
 ALPHA← 0.01 0.03 0.06 0.12 0.22 0.35 0.5
⍝          THIS LOOP INCREMENTS SAMPLE SIZE.
L1:PP←⍴0
 M←⍴0
 KK←⍴0
 ZZ←⍴0
 ZZC←⍴0
 N←N+1
 D←0
 B←0
⍝          COMPUTE CUMULATIVE DIST AND ASSOCIATED STATS.
 P←BINOM N
⍝          THIS LOOP CALCULATES ALPHA VALUES AND APPORXIMATIONS.
L2:B←B+1
 C←ALPHA[B]
 A←(+/(P≤C))
 K←A-1
 KK←KK,K
 PP←PP,P[A]
⍝          COMPUTE NORMAL APPROXIMATION.
 Z←(K-(0.5×N))÷(0.5×(N*0.5))
 ZZ←ZZ,(NORMCDF Z)
⍝          COMPUTE NORMAL APPROXIMATION WITH CONTINUITY CORRECTION.
 ZC←((K+0.5)-(0.5×N))÷(0.5×(N*0.5))
 ZZC←ZZC,(NORMCDF ZC)
 →(B<7)/L2
⍝           COMPUTE ERROR DIFFERENCES.
 ERRZZ←PP-ZZ
 ERRZZC←PP-ZZC
⍝           PRINT OUT TABLE OF VALUES.
 I←0.1
```

```
  J←195
  S←'TEST STAT. VALUE   '
  M←M,'                     ,PROB[K ≤ K̲]; FOR SAMPLE SIZE EQUAL TO ',(2 0 ⍕N),'
  .
  →L3
 L5:J←194
 L3:M←M,[I] ⎕AV[(18⍴197),J,61⍴(8⍴197),J]
  →(D=4)/L4
  H← 1 18 ⍴S
  →(D≠0)/L6
  F←'BG<|  ZZ9   >' ⎕FMT(1 7 ⍴KK)
  F← 1 62 ⍴F
  FS←H,F
  →L7
 L6:FS←⍴0
  H←17⍴S
  FS←FS,H
  C←0
 L9:C←C+1
  F←' |',(7 5 ⍕PP[C])
  FS←FS,F
  →(C<7)/L9
 L7:M←M,[1] FS
  D←D+1
  I←1
  →(N1,N2,N3,L5)[D]
 N1:S←'EXACT C.D.F.       '
  J←198
  →L3
 N2:S←'ERROR; NORMAL      '
  PP←ERRZZ
  →L3
 N3:S←'ERROR; NORM. W/CC '
  PP←ERRZZC
  →L3
 L4:M
  →(N<26)/L1
  ∇


  ∇ SPMTEST;N;ALPHA;B;A;C;AA;P;Z;ZZ;ERRZZ;M;ZZC;D;PP;I;H;F;FS;S;J;TT;T;TTC;
    TC;ERRAVE;ERRTT;ERRTC;ERRZZC;ZC;PC;RHO;KK

 ⍝ THIS PROGRAM GENERATES TABLES OF C.D.F. COMPARISONS FOR SPEARMAN'S R.

 ⍝          SET SAMPLE SIZE AND ALPHA VALUES.
  N←8
  ALPHA← 0.01 0.02 0.03 0.05 0.08 0.13 0.18
 ⍝          THIS LOOP INCREMENTS SAMPLE SIZE.
 L1:PP←⍴0
  M←⍴0
  KK←⍴0
  FS←⍴0
  ZZ←⍴0
  ZZC←⍴0
  TT←⍴0
  TTC←⍴0
  N←N+1
  D←0
  B←0
  AA←6÷N×¯1+N*2
 ⍝          COMPUTE CUMULATIVE DIST AND ASSOCIATED STATS.
 P1:P←PMATSP[N-5;;]
  PC←(P[2;]≠0)/P[2;]
 ⍝          THIS LOOP CALCULATES ALPHA VALUES AND APPORXIMATIONS.
 L2:B←B+1
  C←ALPHA[B]
  A←(+/(PC≤C))
  RHO←P[1;A]
  KK←KK,RHO
  PP←PP,P[2;A]
 ⍝          COMPUTE NORMAL APPROXIMATION.
  Z←RHO×((N-1)*0.5)
  ZZ←ZZ,(1-NORMCDF Z)
 ⍝          COMPUTE NORMAL APPROXIMATION WITH CONTINUITY CORRECTION.
  ZC←(RHO-AA)×((N-1)*0.5)
  ZZC←ZZC,(1-NORMCDF ZC)
 ⍝         COMPUTE STUDENT T APPROXIMATION
  TT←TT,N SPAPROX1 RHO
 ⍝     COMPUTE STUDENT T APPROXIMATION WITH CONTINUITY CORRECTION.
  TTC←TTC,N SPAPROX RHO
  →(B<7)/L2
 ⍝          COMPUTE ERROR DIFFERENCES.
  ERRZZ←PP-ZZ
  ERRZZC←PP-ZZC
  ERRTT←PP-TT
  ERRTC←PP-TTC
 ⍝          PRINT OUT TABLE OF VALUES.
  I←0.1
  J←195
  S←'TEST STAT. VALUE  '
  M←M,'                    ,PROB[R ≥ R̲]; FOR SAMPLE SIZE EQUAL TO ',(2 0 ⍕N),'
  .
 →L3
L5:J←194
L3:M←M,[I] ⎕AV[(18⍴197),J,61⍴(8⍴197),J]
 →(D=6)/L4
 B← 1 18 ⍴S
 →(D=0)/L6
 F←'K4G<| 9.9999 >' ⎕FMT(1 7 ⍴KK)
 F← 1 62 ⍴F
 FS←B,F
 →L7
L6:FS←⍴0
 B←17⍴S
 FS←FS,B
 C←0
L9:C←C+1
 F←' | ',(7 5 ⍕PP[C])
 FS←FS,F
 →(C<7)/L9
L7:M←M,[1] FS
 D←D+1
 I←1
 →(N1,N2,N3,N4,N5,L5)[D]
N1:S←'EXACT C.D.F.      '
 J←198
 →L3
N2:S←'ERROR; NORMAL     '
 PP←ERRZZ
 →L3
N3:S←'ERROR; NORM. W/CC'
 PP←ERRZZC
 →L3
N4:S←'ERROR; T DIST     '
 PP←ERRTT
 →L3
N5:S←'ERROR; T W/CC     '
 PP←ERRTC
 →L3
L4:M
 →(N<10)/L1
 ∇


 ∇ WILTEST;N;ALPHA;B;A;C;T;P;NUM;DEN;Z;ZZ;ERRZZ;M;ZZC;D;PP;NUMC;I;H;F;FS;S
   ;J;TT;T;TTC;TC;NUMC;NUMT;DENOM;DENOMC;ERRAVE;ERRTT;ERRTC;ERRZZC;ZC;ERRT
   CZC;KK

 ⍝ THIS PROGRAM GENERATES TABLES OF C.D.F. COMPARISONS FOR THE WILCOXON
 ⍝ SIGNED-RANK TEST.

 ⍝          SET SAMPLE SIZE AND ALPHA VALUES.
 N←8
 ALPHA← 0.01 0.02 0.03 0.05 0.08 0.13 0.18
 ⍝          THIS LOOP INCREMENTS SAMPLE SIZE.
L1:PP←⍴0
 M←⍴0
 KK←⍴0
 FS←⍴0
 ZZ←⍴0
 ZZC←⍴0
 TT←⍴0
 TTC←⍴0
 N←N+1
 D←0
 B←0
 ⍝        COMPUTE CUMULATIVE DIST AND ASSOCIATED STATS.
 P←WILP N
 ⍝        THIS LOOP CALCULATES ALPHA VALUES AND APPORXIMATIONS.
L2:B←B+1
 C←ALPHA[B]
 A←(+/(P≤C))
 T←A-1
 KK←KK,T
 PP←PP,P[A]
 ⍝        COMPUTE NORMAL APPROXIMATION.
 NUM←T-((N×(N+1))÷4)
 DEN←((N×(N+1)×((2×N)+1))÷24)*0.5
 Z←NUM÷DEN
 ZZ←ZZ,NORMCDF Z
 ⍝        COMPUTE NORMAL APPROXIMATION WITH CONTINUITY CORRECTION.
 NUMC←(T+0.5)-((N×(N+1))÷4)
 ZC←NUMC÷DEN
 ZZC←ZZC,NORMCDF ZC
 ⍝       COMPUTE STUDENT T APPROXIMATION
 DENOM←(((N×(DEN*2))÷(N-1))-((NUM*2)÷(N-1)))*0.5
 T←NUM÷DENOM
 TT←TT,(N-1) TDISTN T
 ⍝    COMPUTE STUDENT T APPROXIMATION WITH CONTINUITY CORRECTION.
 NUMT←(NUM)+0.5
 DENOMC←(((N×(DEN*2))÷(N-1))-((((NUM)+0.5)*2)÷(N-1)))*0.5
 TC←NUMT÷DENOMC
 TTC←TTC.((N-1) TDISTN TC)
 →(B<7)/L2
⍝          COMPUTE ERROR DIFFERENCES.
 ERRZZ←PP-ZZ
 ERRZZC←PP-ZZC
 ERRTT←PP-TT
 ERRTC←PP-TTC
 ERRAVE←PP-((ZZ+TT)÷2)
 ERRTCZC←PP-((ZZC+TTC)÷2)
⍝          PRINT OUT TABLE OF VALUES.
 I←0.1
 J←195
 S←'TEST STAT. VALUE  '
 M←M,'                        PROB[W ≤ W̲]; FOR SAMPLE SIZE EQUAL TO ',(2 0 ⍕N),'
:
 →L3
L5:J←194
L3:M←M,[I] ⎕AV[(18⍴197),J,61⍴(8⍴197),J]
 →(D=8)/L4
 H← 1 18 ⍴S
 →(D=0)/L6
 F←'BG<|   ZZ9    >' ⎕FMT(1 7 ⍴KK)
 F← 1 62 ⍴F
 FS←H,F
 →L7
L6:FS←⍴0
 H←17⍴S
 FS←FS,H
 C←0
L9:C←C+1
 F←' | ',(7 5 ⍕PP[C])
 FS←FS,F
 →(C<7)/L9
L7:M←M,[1] FS
 D←D+1
 I←1
 →(N1,N2,N3,N4,N5,N6,N7,L5)[D]
N1:S←'EXACT C.D.F.      '
 J←198
 →L3
N2:S←'ERROR; NORMAL     '
 PP←ERRZZ
 →L3
N3:S←'ERROR; NORM. W/CC'
 PP←ERRZZC
 →L3
N4:S←'ERROR; T DIST     '
 PP←ERRTT
 →L3
N5:S←'ERROR; T W/CC     '
 PP←ERRTC
 →L3
N6:S←'ERROR; AVE T/Z    '
 PP←ERRAVE
 →L3
N7:S←'ERROR; AVE TC/ZC '
 PP←ERRTCZC
 →L3
L4:M
 →(N<10)/L1
∇
```

INITIAL DISTRIBUTION LIST

No. Copies

1. Defense Technical Information Center 2
Cameron Station
Alexandria, Virginia 22304-6145

2. Library, Code 0142 2
Naval Postgraduate School
Monterey, California 93943-5002

3. Superintendent, Code 53Jy 1
Attn: Prof. T. Jayachandran
Naval Postgraduate School
Monterey, California 93943

4. Superintendent, Code 55Bm 1
Attn: Prof. D. Barr
Naval Postgraduate School
Monterey, California 93943

5. Superintendent, Code 55La 1
Attn: Prof. H. Larson
Naval Postgraduate School
Monterey, California 93943

6. Superintendent, Code 55Re 1
Attn: Prof. R. Read
Naval Postgraduate School
Monterey, California 93943

6. Superintendent, Code 55Rh 1
Attn: Prof. R. Richards
Naval Postgraduate School
Monterey, California 93943

7. Superintendent, Code 55Wd 1
Attn: Prof. K. Wood
Naval Postgraduate School
Monterey, California 93943

7. Commanding Officer 2
Naval Special Warfare Unit - Four
Box 3400
FPO Miami, Florida 34051

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93943-5002

Thesis
V1237 Vagts
c.1 An APL workplace for
conducting nonparametric
statistical inference.

Thesis
V1237 Vagts
c.1 An APL workplace for
conducting nonparametric
statistical inference.

thesV1237
An APL workspace for conducting nonparam
3 2768 000 73708 4
DUDLEY KNOX LIBRARY